

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE CIENCIAS MATEMATICAS

E.A.P. DE ESTADISTICAS

**“COMPARACIÓN DE LOS MODELOS LOGIT Y
PROBIT DEL ANÁLISIS MULTINIVEL, EN EL
ESTUDIO DEL RENDIMIENTO ESCOLAR”**

TESIS

**Para optar el Título Profesional de
Licenciado en Estadística.**

AUTOR

Víctor Hugo Ucedo Silva

LIMA – PERÚ

2013

COMPARACIÓN DE LOS MODELOS LOGIT Y PROBIT DEL ANÁLISIS MULTINIVEL

VÍCTOR HUGO UCEDO SILVA

Tesis presentada a consideración del Cuerpo Docente de la Facultad de Ciencias Matemáticas, de la Universidad Nacional Mayor de San Marcos, como parte de los requisitos para obtener el Título Profesional de Licenciado en Estadística.

Aprobada por:

Mg. Olga Lidia Solano Dávila

Mg. Ana María Cárdenas Rojas

Lic. Rosa Fátima Medina Merino

Lima – Perú
Noviembre – 2013

FICHA CATALOGRÁFICA

UCEDO SILVA, VÍCTOR HUGO

Comparación de los modelos LOGIT y PROBIT del
Análisis Multinivel, (Lima, 2013).

x, 126 p., 21 cm x 29.7cm, (UNMSM, Licenciado en
Estadística, 2013).

Tesis, Universidad Nacional Mayor de San Marcos,
Facultad de Ciencias Matemáticas. Estadística. UNMSM /
F de CM. Comparación de los modelos LOGIT y PROBIT
del Análisis Multinivel.

La falta más grave es no tener conciencia de nuestra propia falta.

*A Dios, por su gran amor,
que nos llena e inspira a amar.*

*Y a quienes amamos con
todas nuestras fuerzas, por
estar a nuestro lado, mi esposa
Geraldine e hija Luciana.*

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a mi asesora Lic. Fátima Medina Merino. Por su apoyo, voluntad y dedicación en el desarrollo de la presente tesis.

Mi agradecimiento a todas mis profesoras de la Facultad de Ciencias Matemáticas de la UNMSM por sus enseñanzas.

Finalmente mi agradecimiento a la UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS, por permitir sentirme orgulloso de llamarla Alma Mater.

RESUMEN

COMPARACIÓN DE LOS MODELOS LOGIT Y PROBIT DEL ANÁLISIS MULTINIVEL

VÍCTOR HUGO UCEDO SILVA

NOVIEMBRE – 2013

Orientador : Lic. Fátima Medina Merino
Título Obtenido : Licenciado en Estadística

En esta investigación se desea comparar dos modelos Logit y Probit del Análisis Multinivel, la comparación se realizó a través de las Medidas de Calidad para el Ajuste Global, el estadístico de Wald y la diferencia de las razones de verosimilitudes. Para el análisis se empleó una Base de Datos del Programa Internacional para la Evaluación de Estudiantes (PISA) del año 2009. Asimismo se presentó las similitudes y diferencias en el uso de cada modelo.

Palabras Claves: Análisis Multinivel, Logit, Probit, Correlación Intraclase.

ABSTRACT

COMPARISON OF MODELS LOGIT AND PROBIT OF MULTILEVEL ANALYSIS

VÍCTOR HUGO UCEDO SILVA

NOVEMBER – 2013

Two methods of multilevel analysis are compared: LOGIT and PROBIT. The comparison was made through his quality measures for global fit, Wald statistic and the difference of likelihood ratios. For the analysis used a database of Program for International Student Assessment (PISA) 2009. Also it presented the similarities and differences in the use of each model.

Keys Words: *Multilevel Analysis, Logit, Probit, Intraclass Correlation.*

ÍNDICE

INTRODUCCIÓN.....	11
Capítulo I: Análisis Multinivel	20
1.1 Introducción.....	20
1.2 Definición	20
1.2.1 Estructura de los Datos	22
1.2.2 Terminología	25
1.2.3 Tipos de Falacia	26
1.3 Modelos de Regresión Clásicos	27
1.3.1 Modelo de regresión de un sólo nivel con datos desagregados	27
1.3.2 Modelo de regresión de un sólo nivel con datos agregados	27
1.3.3 Ecuaciones de Regresión Separadas	28
1.3.4 Modelo de regresión simple con variables individuales y contextuales	31
1.4 Estimación de parámetros en Modelos Multinivel	32
1.5 Coeficiente de Correlación Intraclass (CCI)	33
1.6 Funciones de enlace para variables de respuesta binaria	35
1.7 Validación del Modelo	38
1.7.1 Evaluación de la Capacidad predictiva del Modelo	40
1.7.2 Prueba de hipótesis individual de las variables	41
1.8 Medidas de Calidad para el Ajuste Global de los Modelos	43
1.8.1 Criterio de Información de Akaike AIC	44
1.8.2 Criterio de Información de Akaike corregido AICC	44
1.8.3 Criterio de Información de Akaike consistente CAIC	44
1.8.4 Criterio de Información Bayesiano BIC	44

1.9 Ventajas	46
1.10 Desventajas	46
 Capítulo II: Modelo LOGIT	 47
2.1 Introducción.....	47
2.2 Definición	48
2.3 Función de Enlace LOGIT	49
2.3.1 Modelo Nulo	51
2.3.2 Modelo con un criterio	51
2.3.3 Modelo con intercepto aleatorio y variables contextuales	51
2.3.4 Modelo Modelo con intercepto y pendientes aleatorias	52
2.3.5 Coeficiente de Correlación Intraclase	52
2.4 Ventajas del modelo LOGIT	53
2.5 Desventajas del modelo LOGIT	53
 Capítulo III: Modelo PROBIT	 54
3.1 Introducción.....	54
3.2 Definición	55
3.3 Función de Enlace PROBIT	56
3.3.1 Modelo Nulo	58
3.3.2 Modelo con un criterio	58
3.3.3 Modelo con intercepto aleatorio y variables contextuales	58
3.3.4 Modelo Modelo con intercepto y pendientes aleatorias	59
3.3.5 Coeficiente de Correlación Intraclase	59
3.4 Ventajas del modelo PROBIT	60
3.5 Desventajas del modelo PROBIT	60
 Capítulo IV: Resultados	 63
4.1 Variables e indicadores	67
4.2 Base de Datos: PISA	72
4.3 Análisis Multinivel en LISREL	76

4.3.1 Modelo Logit	78
4.3.2 Modelo Final Logit	88
4.3.3 Modelo Probit	94
4.3.4 Modelo Final Probit	102
Capítulo V: Conclusiones	108
Capítulo VI: Limitaciones y Recomendaciones	114
BIBLIOGRAFÍA	115
ANEXOS	121
Anexo 1: Modelo 1 - LOGIT	122
Anexo 2: Modelo 2 - LOGIT	125
Anexo 3: Modelo 3 - LOGIT	130
Anexo 4: Modelo 1 - PROBIT	134
Anexo 5: Modelo 2 - PROBIT	138
Anexo 6: Modelo 3 - PROBIT	141

INTRODUCCIÓN

Los Modelos de Análisis Multinivel constituyen la metodología de análisis más adecuada para explicar datos jerarquizados, lo que la convierte en una herramienta imprescindible para la investigación. Además de mejorar la calidad de los resultados respecto a sus estimaciones, posibilita realizar análisis novedosos, tales como estimar la aportación de cada nivel de análisis o las interacciones entre variables de distintos niveles.

Entre los modelos de **Análisis multinivel** más conocidos están el **Logit** y el **Probit**, utilizados para describir la relación existente entre una variable respuesta de tipo dicotómica y una o más variables explicativas. A través de sus estimaciones puntuales e intervalares de los parámetros se calcularon medidas de significancia y parsimonia del modelo.

En conclusión se realizó una comparación de los modelos obtenidos mediante el Análisis Multinivel, presentando las similitudes y diferencias de estos modelos. Además se calculó los resultados de los estadísticos de las Medidas de Calidad para el Ajuste Global, el estadístico de Wald (a través del estadístico Z) y la Diferencia de las Razones de Verosimilitudes, los cuales fueron de utilidad para determinar el modelo más adecuado. Para el análisis se empleará una Base de Datos del Programa Internacional para la Evaluación de Estudiantes (PISA) del año 2009.

DEFINICIÓN DEL PROBLEMA:

En muchas oportunidades para poder llegar a la unidad de análisis y obtener datos del mismo, se ha tenido que recurrir a un diseño muestral complejo ya que se han pasado por varias etapas, es decir nuestros datos están bajo una

estructura compleja, sin embargo algunas técnicas estadísticas tradicionales no toman en cuenta dicha estructura del cual provienen los datos, sino más bien, por el contrario asumen que los datos han sido tomados bajo un muestreo aleatorio simple. A partir de ello nace la necesidad de buscar, conocer, y aprender técnicas que tomen en cuenta esta estructura compleja, para poder aplicarla en el tema de investigación y especialmente bajo el contexto de estudio que estemos realizando. Una propuesta necesaria e interesante fue revisar las diferentes literaturas que aplican técnicas estadísticas que toman en cuenta el contexto de los datos como son los Modelos del Análisis Multinivel.

OBJETIVOS GENERAL:

COMPARAR LOS MODELOS LOGIT Y PROBIT DEL ANÁLISIS MULTINIVEL, EN EL ESTUDIO DEL RENDIMIENTO ESCOLAR.

OBJETIVOS ESPECÍFICOS:

- Conocer las similitudes y diferencias entre los modelos Logit y Probit del análisis multinivel.
- Validar los modelos Logit y Probit del análisis multinivel.
- Conocer las ventajas y desventajas en la aplicación de cada uno de los modelos.

JUSTIFICACIÓN:

La importancia de este problema se basa en aplicar los modelos estadísticos adecuados, y evitar encontrar relaciones entre las variables de manera incorrecta ya que no necesariamente dos variables están relacionada a

nivel individual debido a que pueden estar relacionadas a un nivel mayor, esto es conocido como ***falacia***, el cual se clasifica en atomista y ecológica. La falacia atomista surge cuando se realizan inferencias sobre la variabilidad entre los grupos a partir de datos individuales, y en sentido contrario se conoce como falacia ecológica.

Resumen

En el capítulo I se desarrolló el Modelo de Análisis Multinivel. Los Modelos Multinivel constituyen una propuesta realista del análisis de los datos, debido a que los diseños muestrales que se emplean actualmente son complejos.

En el capítulo II se desarrolló el Modelo LOGIT. La regresión Logit utiliza una función de distribución acumulativa (FDA) de tipo logístico. La regresión logística se basa en la suposición de que la variable dependiente categórica refleja una variable subyacente cualitativa (éxito, fracaso) y deduce la función de enlace a partir de la distribución binomial.

En el capítulo III se desarrolló el Modelo PROBIT, Para la regresión Probit se utiliza una FDA de distribución normal, (modelo Probit en muchas ocasiones se le llame también "Modelo Normit"). La regresión Probit asume que la variable dependiente categórica refleja una distribución subyacente cuantitativa la cual ha sido recategorizada de modo que se convierte en una variable binaria y deduce la función de enlace de la distribución normal estándar acumulada.

En el capítulo IV se presenta los resultados del Modelo de Análisis Multinivel, con la Función de enlace Probit, (Probit Multinivel) y la Función de enlace Logit (Logit Multinivel) con ayuda del software LISREL. Además se plantean de manera formal los modelos para cada caso y se obtienen la razón de

verosimilitud, entre otros indicadores. Por otro lado se obtiene el estadístico de Wald, y tablas de probabilidad para cada uno de los modelos, calculados en el desarrollo de la presente tesis. Para el análisis se empleó la Base de Datos del Programa Internacional para la Evaluación de Estudiantes (PISA) del año 2009.

En las siguientes líneas presentamos algunos antecedentes de investigaciones internacionales en las cuales se tiene como referencia aplicaciones con las técnicas del análisis multinivel.

Montero, E., Villalobos, J., Valverde, A. (2007). Propusieron un modelo para predecir las calificaciones finales en los cursos de carrera a través de un Análisis Multinivel de dos niveles, cuya muestra estuvo conformada por 848 estudiantes los cuales pertenecían a diferentes áreas, el tipo de muestreo que utilizaron fue complejo, (muestreo aleatorio estratificado de conglomerados completos) en el cual tomaron los grupos como conglomerados y las áreas académicas como estratos, buscando que la muestra sea representativa por las áreas académicas. Entre sus principales factores estudiados se conformaron los aspectos institucionales, sociodemográficos, psicosociales, y pedagógicos. Específicamente en el modelo que utilizaron de regresión multinivel sus dos niveles se detallan de la siguiente manera: el primer nivel estuvo conformado por las variables del estudiante, y en el segundo nivel las variables del profesor y del curso. Entre sus indicadores estadísticos del modelo mencionan al coeficiente de correlación intraclase el cual afirman es provocado por la conglomeración de los individuos en cursos, en centros educativos, etc. En su investigación mencionan con detalle las variables que consideraron en el modelo multinivel desde la edad, el colegio de procedencia, si el estudiante trabaja, entre otros, hasta los índices de satisfacción del estudiante con sus respectivos rangos.^[1]

Álvarez, D. y otros. (2006 - 2007), realizaron una investigación multinivel sobre el impacto de las características ambientales de la zona de residencia en la salud mental de población autóctona e inmigrante. El objetivo de este artículo es

describir el planteamiento metodológico de la investigación, el trabajo de campo, las tasas de respuesta correspondientes y discutir el diseño metodológico y las dificultades derivadas de su puesta en práctica. Los datos individuales se obtuvieron aplicando un cuestionario estructurado de aproximadamente 40 minutos, mediante entrevista domiciliaria a personas españolas y ecuatorianas de 18 a 55 años. La muestra estimada fue de 1186 personas equiparada por sexo y nacionalidad obtenida aleatoriamente de los padrones municipales de 33 áreas (municipios o barrios) de Madrid, Alicante, Almería y Murcia, seleccionadas según criterios de densidades étnicas y socioeconómicas. Por ello se diseñó una investigación con dos niveles de análisis: a) los individuos, mediante encuesta de base poblacional con datos obtenidos por visita y entrevista domiciliaria y b) las áreas de residencia, mediante datos obtenidos de fuentes secundarias.^[2]

Cervini, R. (2006), presenta la investigación denominada “Trabajo infantil y progreso aprendizaje en la educación básica” para ello utilizó un modelo de Análisis multinivel de “valor agregado”, en convenio entre la Dirección General de Cultura y educación de Buenos Aires y la Universidad Nacional de Quilmes. En la investigación mencionada presentan como objetivo explorar el efecto del trabajo infantil sobre el progreso de aprendizaje del alumno en Matemática de la Educación básica, entre otros temas, como conocer el efecto dentro y entre las escuelas. La muestra fue de 5268 alumnos en 88 escuelas, entre las variables se consideró la situación laboral, antecedentes socioeconómicos y demográficos del alumno y la composición académica y socioeconómica de la escuela, además consideraron como variable dependiente al logro final que fue el puntaje obtenido por el alumno (se consideró una cohorte de alumnos desde el 7mo al 9no, 2003) en la prueba de matemática, el análisis multinivel tuvo dos niveles, como unidad del primer nivel se consideró al alumno, y para el nivel dos, la escuela. En este artículo se presentan los modelos del análisis multinivel, entre sus conclusiones de las cuales llama la atención es que el trabajo precoz tiene un efecto negativo sobre el progreso en el aprendizaje escolar de Matemática. Además que las tasas

de progreso relativo son significativamente inferiores en los alumnos que trabajan respecto a los que no trabajan.^[3]

Calero, J., Escardíbul, J., Waisgrais, S., Mediavilla, M. (2003). Analizaron la importancia del entorno socioeconómico y cultural de los individuos sobre su rendimiento académico. Para ello, utilizaron los datos originados en PISA 2003, una evaluación de conocimientos que ha desarrollado la OCDE, en dicho año y para diversos países, en escuelas con alumnos y alumnas de 15 años de edad. Respecto al entorno de los estudiantes que se presentaron, enfatizaron en diversos elementos vinculados tanto con la familia como con la escuela. Entre los primeros destacan la clase social, el capital humano de los padres (educación y cultura) y la nacionalidad; entre los segundos, el tipo de centro escolar (público o privado) y las características socioeconómicas de éste: clima educativo (nivel medio de estudios de los padres de los alumnos), perfil social y grado de presencia de los inmigrantes. Con el modelo multinivel, se logra conocer que parte de los efectos de las variables independientes se deben a los individuos y que parte a las escuelas a las que asisten.^[4]

De Onis, M., Frongillo, E., Blossner, M. (2001), analizaron de forma normalizada datos transversales de 241 encuestas representativas de países a fin de obtener resultados comparables en lo que respecta a la estatura baja para la edad (retraso del crecimiento) se efectuó una modelización multinivel para estimar las tendencias regionales y mundiales entre 1980 y 2005. En el análisis estadístico se efectuó una modelización multinivel, consistente en una generalización de la regresión estándar mediante el uso de niveles y unidades múltiples, para estimar las tendencias regionales y mundiales del retraso del crecimiento. El conjunto de datos que establecieron tenía tres niveles: regiones, países, y año de realización de la encuesta.

Los resultados que obtuvieron son muy interesantes, en este resumen presentamos parte de ellos, seguidamente: Encontraron que la prevalencia del

retraso del crecimiento ha disminuido en los países en desarrollo de un 47% en 1980 a un 33% en 2000 (es decir, en 40 millones de personas), aunque el progreso ha sido desigual en las diferentes regiones. Los casos de retraso del crecimiento han aumentado en África oriental, pero han disminuido en Asia sudoriental, meridional y central, América del Sur (Bolivia, Brasil, Chile, Colombia, Guyana, Perú, Uruguay, Venezuela), en Perú hubo una disminución en la Variación anual de 0.66%; África septentrional y el Caribe muestran una ligera mejora; y en el África occidental y Centroamérica los progresos han sido muy escasos.

Pese a la disminución global del retraso del crecimiento en los países en desarrollo, la malnutrición infantil sigue siendo un serio problema de salud pública en esos países. En algunos de ellos las tasas de retraso del crecimiento están aumentando, mientras que en otros se mantienen elevadas de manera inquietante. La información que se presentó proporciona datos de referencia útiles para valorar los progresos permitir identificar los países y regiones que necesitan intervenciones destinadas a toda la población. Las perspectivas aplicadas para reducir la malnutrición infantil deberían basarse en los programas y políticas nutricionales satisfactorias. ^[5]

En el Perú también se han realizado algunas investigaciones en las cuales se aplicó el análisis multinivel.

Benavides, M., Caro, D., Espinosa, G., Miranda, L., Tam, M., Zacharías, D., Zambrano, G. (2004). El Equipo de análisis de la Unidad de Medición de la Calidad Educativa (**UMC**), realizaron una investigación que fue llamada “Factores Asociados al Rendimiento Estudiantil”, en la que utilizaron los Modelos Multinivel. En su informe presentan los principales resultados a partir de los modelos de factores asociados al rendimiento de los estudiantes de cuarto de primaria y cuarto de secundaria, evaluados en el área de matemática y comunicación en el año

2001. Realizaron el análisis a partir de dos niveles, alumnos y escuelas. El tamaño de muestra que tomaron para la parte multivariada fue de 5729 estudiantes que corresponden a 625 centros educativos en cuarto de primaria. Mientras que en el cuarto de secundaria se tomó un tamaño de muestra de 13416 alumnos que correspondía a 569 centros educativos. Hubo una reducción en los tamaños de muestra comparados con los tomados en la parte descriptiva especialmente en el caso del cuarto de primaria, debido a que no fue aplicado en su totalidad el cuestionario a los padres de los estudiantes. Para el desarrollo del análisis multinivel, las variables escolares fueron imputadas con la finalidad de no tener pérdida de información. A nivel individual (intraescolares) entre las dimensiones consideradas se encuentran las características familiares, del alumno. Respecto a las características familiares se tienen el número de libros, nivel educativo de los padres, sus expectativas entre otras. Y en las características del alumno se tienen variables individuales como su género, edad, lengua materna, tiempo del hogar a la escuela, si trabaja durante los días de escuela o los fines de semana, la edad que ingreso a primaria, entre otras. A nivel escolar (entre escuelas) se cuenta con características de ubicación, ya sea en costa o sierra, en el área urbana o rural, también se tienen características institucionales, de composición, equipamiento, procesos educativos como el manejo de docente, carga laboral, preparación de sus clases y otras variables adicionales. ^[6]

PERSPECTIVA EN EL AMBITO DEL PAIS:

Hace algunos años el Perú ha participado a nivel internacional en área de educación, específicamente nuestros alumnos de 15 años han sido evaluados en comprensión de lectura (PISA 2009), a pesar que los resultados de la evaluación no fueron alentadores, esto nos motiva a no darnos por vencidos, para tal fin debemos buscar las formas de tomar las medidas adecuadas para mejorar, desde el enfoque de la estadística buscar los instrumentos, y las técnicas que en conjunto nos den el método adecuado. En la metodología aplicada en el estudio mencionado se utilizaron los modelos de análisis multinivel, el cual les garantiza

resultados coherentes debido a la estructura compleja del sector.

Desde esta perspectiva, estos modelos ayudarían a proponer estudios continuos, una iniciativa propia que se ajuste a nuestra realidad, a nuestros contextos, que faciliten a conocer nuestras debilidades y fortalezas en el país, que estén más allá del periodo de un gobierno, y que trasciendan a través de los diferentes periodos, no basta un objetivo común en un determinado momento, se espera que trascienda en el tiempo, pero no solo el sector de educación, también en el sector salud, entre otros sectores, debido a sus propios contextos, ya sea a un nivel micro y/o macro, para así tener mediciones estructuradas de manera constante.

CAPÍTULO I:

Análisis Multinivel

1.1 Introducción

Los Modelos Jerárquicos o Modelos Multinivel constituyen una propuesta conceptualmente más realista del análisis de los datos, debido a que los diseños muestrales que se emplean actualmente son complejos, es decir son distintos al diseño de un muestreo aleatorio simple. Además simultáneamente, en un mismo modelo se consideran variables medidas a nivel micro y macro, así como las interacciones entre variables medidas en contextos diferentes.

1.2 Definición

El Análisis Multinivel también es conocido como Análisis Jerárquico, de Componentes de Varianza, de Efectos Aleatorios o de Coeficientes Aleatorios.

El Análisis Multinivel nos permite estudiar el efecto simultáneo de características individuales como contextuales además de sus interacciones. Es considerada como una metodología de análisis para conjuntos de datos en los cuales sus patrones de variabilidad son complejos debido a que provienen de una estructura jerárquica como son los diseños muestrales multietápicos o por relaciones también complejas entre las variables de estudio como los modelos estructurales.

Un modelo multinivel es capaz de tratar los contextos particulares como si fueran muestras aleatorias extraídas de una población mucho más general de contextos (unidades de nivel macro).

Esta técnica permite realizar inferencias acerca de la variación entre todos los contextos en una población a partir de una muestra aleatoria de contextos y de individuos.

El análisis multinivel debe ser aplicado en diseños multietápicos, estudios de panel, modelos de medidas repetidas con tiempos desigualmente espaciados, modelos multinivel multivariados, análisis de series de tiempo multinivel, modelos de regresión lineal múltiple multinivel y modelos de regresión lineal generalizada multinivel.

Andréu, J. (2011). El análisis multinivel es una respuesta a la necesidad de analizar la relación entre los individuos y los diversos contextos en los que se desenvuelven. Las hipótesis de partida de estos modelos nos dicen que los individuos pertenecientes a un mismo contexto tenderán a ser más similares en su comportamiento entre sí, que respecto a su pertenencia a distintos contextos.^[7]

El objetivo fundamental del análisis multinivel, es modelar estadísticamente la influencia de variables contextuales sobre las actitudes o los comportamientos medidos a nivel individual

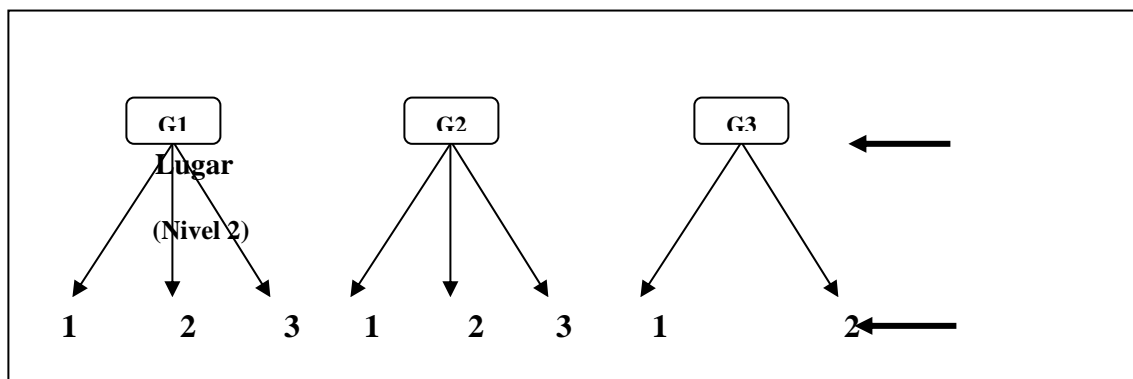
1.2.1 Estructura de los datos

En base a la revisión que se ha realizado respecto a las definiciones principales del análisis multinivel, los datos provienen de una estructura jerárquica esto quiere decir que las observaciones, los individuos están anidados dentro de subgrupos, a su vez estos subgrupos dentro de grupos, y así esta secuencia podría continuar. Es por ello que podemos tener diseños muestrales complejos (como los diseños multietápicos que puede tener una estructura de dos niveles o bien de tres niveles, etc.). A veces en dicha estructura puede intervenir el tiempo como en los diseños de medidas repetidas, inclusive los diseños de tipo panel. Seguidamente presentamos de manera gráfica estos casos:

1.2.1.1 Diseño muestral de dos etapas

En esta estructura se considera al individuo como unidad elemental para el nivel 1 y el lugar como elemento principal del nivel 2, así por ejemplo: si se desea analizar el rendimiento académico en los alumnos (nivel 1) sabemos que estos se encuentran agrupados (anidados) dentro de los colegios (nivel 2) entonces sería interesante analizar el rendimiento académico dentro de los colegios y entre los colegios.

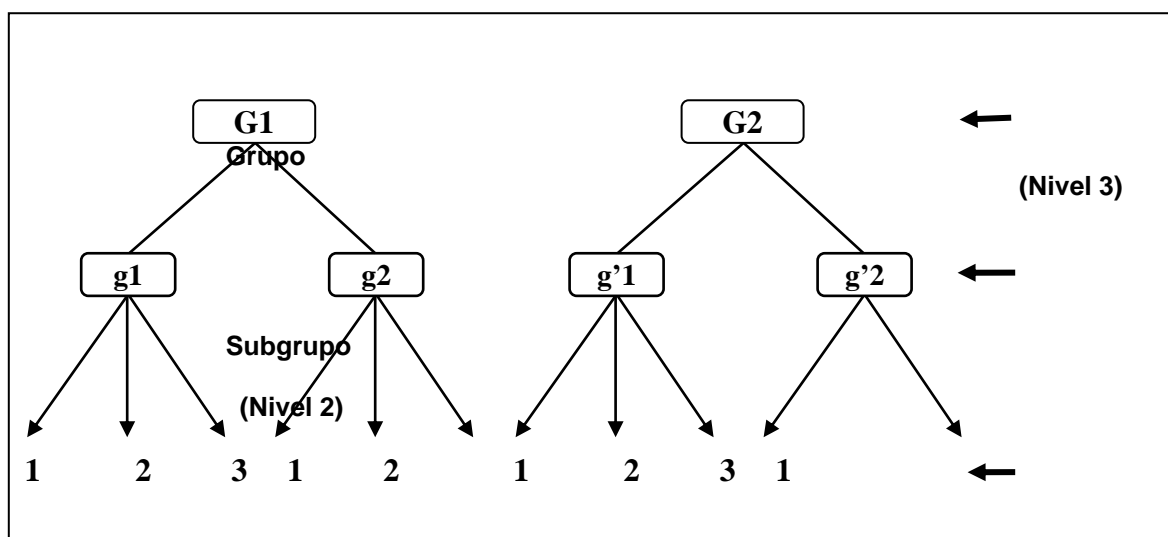
Figura N° I.1. ESTRUCTURA EN DOS NIVELES



1.2.1.2 *Diseño muestral de tres etapas*

En esta estructura se considera al individuo como unidad elemental para el nivel 1 el subgrupo como elemento principal del nivel 2 y el grupo como elemento principal del nivel 3, así por ejemplo: si se desea analizar el rendimiento académico en los alumnos (nivel 1) sabemos que estos se encuentran agrupados (anidados) dentro de diferentes clases (nivel 2) y a su vez estas clases se encuentran ubicadas dentro de un grupo mayor que es el colegio (nivel 3). Esto dependerá básicamente de los objetivos de la investigación y de la estructura de nuestros datos, por ejemplo supongamos que un investigador tiene como unidades de estudio: alumno, docente y colegio se mantiene la misma estructura de tres niveles pero estaríamos ahora ingresando al docente como la unidad de nivel 2 a diferencia del caso anterior.

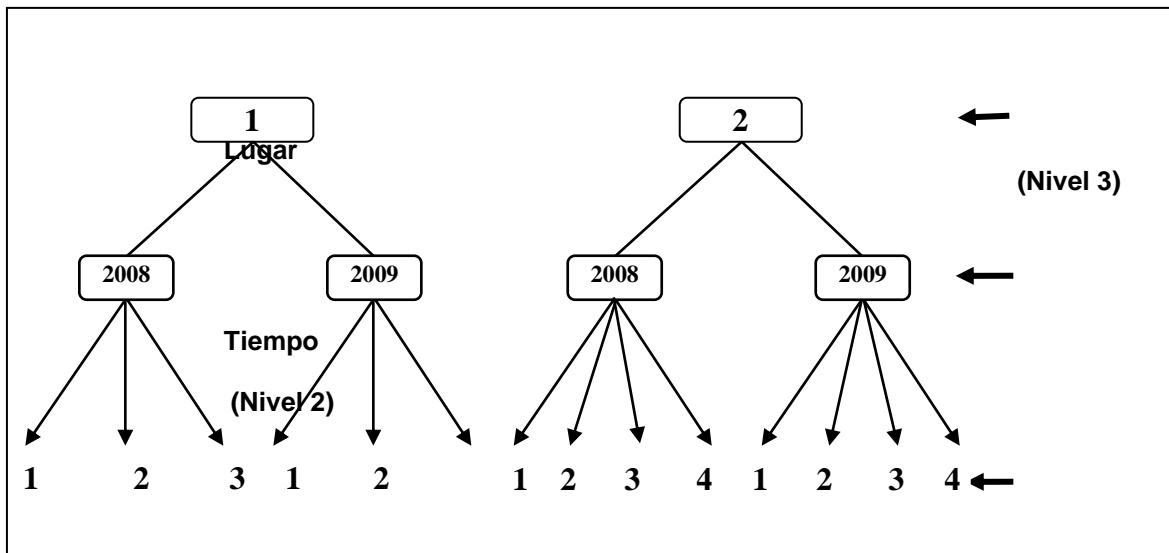
Figura N° I.2. ESTRUCTURA EN TRES NIVELES



1.2.1.3 *Diseño de medidas repetidas*

En el presente diagrama se presenta una estructura de tres niveles, por ejemplo consideremos que en un determinado colegio (nivel 3) se tomo dos mediciones en el año 2008 y 2009 (nivel 2) a una muestra de estudiantes (nivel 1), ahora esto lo extendemos para un conjunto mayor de colegios.

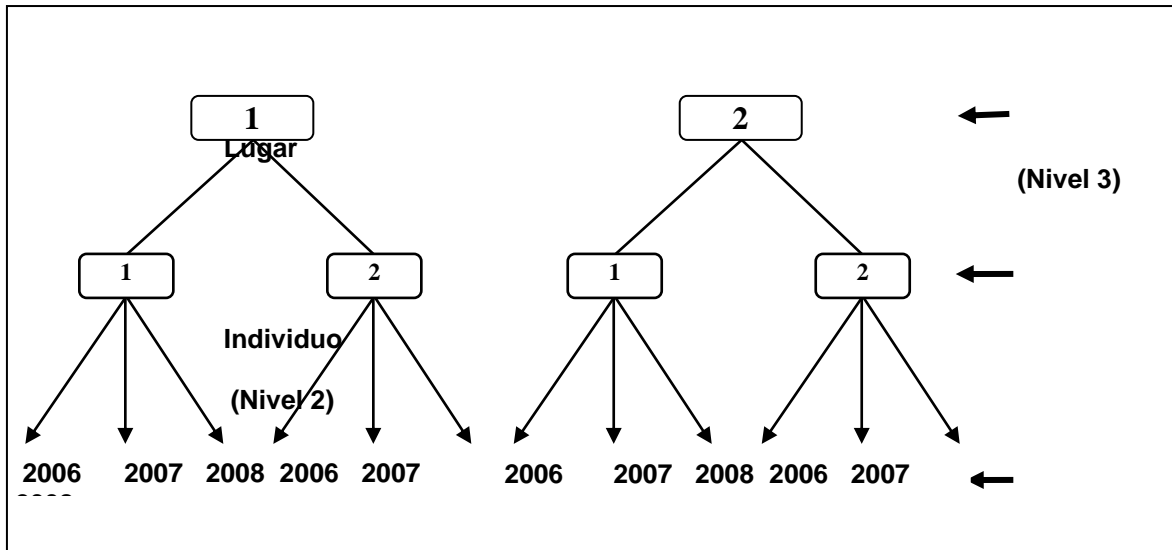
Figura N° I.3. ESTRUCTURA DE MEDIDAS REPETIDAS



1.2.1.4 *Diseño de Panel*

Aquí se toman mediciones en diferentes momentos en el tiempo, del año 2006 al 2008 que pueden haber sido mediciones de un grupo de estudiantes de un determinado profesor que este a su vez se encuentra dentro de un colegio.

Figura N° I.4. ESTRUCTURA DE PANEL



1.2.2 Terminología

1.2.2.1 Nivel: es la estructura o etapas con las que se trabaja en el diseño de la investigación.

Caso de dos niveles:

El cuadro muestra en cada columna términos similares para cada nivel que son utilizados en las diferentes fuentes bibliográficas, se entiende así por ejemplo que las Micro unidades se encuentran dentro de las Macro unidades.

Unidades micro-nivel	→	Unidades macro-nivel
Micro unidades	→	Macro unidades
Unidades primarias	→	Unidades secundarias

Unidades elementales →	Conglomerados
Unidades de nivel 1 →	Unidades de nivel 2
Unidades composicionales →	Unidades contextuales

1.2.2.2 Efecto Fijo: asume que los datos provienen de poblaciones normales las cuales podrían diferir únicamente en sus medias.

1.2.2.3 Efecto Aleatorio: asume que los datos describen una jerarquía de diferentes poblaciones cuyas diferencias quedan restringidas por la jerarquía. Ejemplo: El experimentador ha aprendido y ha considerado en el experimento sólo tres de muchos más métodos posibles, el método de enseñanza es un factor aleatorio en el experimento.

1.2.2.4 Efectos contextuales: se refiere a los fenómenos bajo estudio que operan a nivel macro.

1.2.2.5 Efectos composicionales: se refiere a los fenómenos bajo estudio que operan a nivel micro.

1.2.2.6 Interacción: es cuando la asociación entre dos variables varía según los diferentes niveles de otra u otras variables.

1.2.3 Tipos de Falacia

Las técnicas tradicionales de análisis de datos operan a un sólo nivel, por lo tanto el investigador se ve obligado a decidir en qué nivel va a realizar el análisis:

- Por agregación de los datos a macro unidades calculando tasas, promedios, totales y realizando el análisis a ese nivel, lo cual puede conducir a una “Falacia Ecológica”.
- Por desagregación de los datos a micro unidades, lo cual incrementa el número de observaciones e ignora la variabilidad entre macro unidades. Puede conducir a una “Falacia atomista”.

1.3 MODELOS DE REGRESIÓN CLÁSICOS

1.3.1 Modelo de regresión de un sólo nivel con datos desagregados

Este modelo ignora la variable de nivel 2 y considera todos los individuos como si fueran observadas en una misma población.

$$y_{i1} = \beta_{01} + \beta_{11}X_{i1} + \varepsilon_{i1} ; \quad i = 1, \dots, n ; \quad \varepsilon \sim N(0, \sigma_e^2) \dots \dots (I. 5)$$

Donde:

y_{i1} = i – ésima observación de la variable dependiente del nivel 1

β_{j1} = Coeficiente de regresión en el nivel 1.

X_{i1} = i – ésima observación de la variable independiente del nivel 1

ε_{i1} = Error aleatorio en el nivel 1 ; $i = 1, \dots, n$.

σ_e^2 = Varianza del error aleatorio en el nivel 1.

Aquí se estaría asumiendo que la ocurrencia de la enfermedad no difiere en las comunidades

1.3.2 Modelo de regresión de un sólo nivel con datos agregados

Este modelo ignora la variable de nivel 1 y considera todos los elementos del nivel 2 como si fueran observadas en una misma población.

$$y_j = \gamma_0 + \gamma_1 X_j + \mu_j ; \quad j = 1, \dots, J ; \quad \mu \sim N(0, \sigma_\mu^2) \dots \dots (I.6)$$

Donde:

y_j = j – ésima observación de la variable dependiente del nivel 2.

γ_i = Coeficiente de regresión en el nivel 2.

X_j = j – ésima observación de la variable independiente del nivel 2.

μ_j = Error aleatorio en el nivel 2 ; $j = 1, \dots, n$

σ_μ^2 = Varianza del error aleatorio en el nivel 2

1.3.3 Ecuaciones de Regresión Separadas

$$y_{i1} = \beta_{01} + \beta_{11}X_{i1} + \varepsilon_{i1}$$

$$y_{i2} = \beta_{02} + \beta_{12}X_{i2} + \varepsilon_{i2}$$

.....

$$y_{iJ} = \beta_{0J} + \beta_{1J}X_{iJ} + \varepsilon_{iJ}$$

Aquí las comparaciones tendrán que ser hechas observando los coeficientes de cada uno de los modelos ajustados y se comparan entre sí, (la mayoría de las veces es una verificación no estadística).

1.3.3.1 Modelos con interacción

$$y_i = \gamma_{00} + \gamma_{10}X_{1i} + \gamma_{01}Z_{1j} + \gamma_{11}X_{1i}Z_{1j} + \varepsilon_i \dots \dots \dots (I. 7)$$

Reagrupando términos tenemos:

$$y_i = (\gamma_{00} + \gamma_{01}Z_{1j}) + (\gamma_{10} + \gamma_{11}Z_{1j})X_{1i} + \varepsilon_i$$

Haciendo:

$$\beta_0 = \gamma_{00} + \gamma_{01}Z_{1j}$$

$$\beta_1 = \gamma_{10} + \gamma_{11}Z_{1j}$$

Se tiene:

$$y_i = \beta_0 + \beta_1X_{1i} + \varepsilon_i ; \quad i = 1, 2, \dots, n ; \quad \varepsilon \sim N(0, \sigma_e^2) \dots \dots (I. 8)$$

Donde:

y_i = i – ésima observación de la variable dependiente del nivel 1

γ_{kj} = Coeficiente de regresión en el nivel 2

X_{1i} = i – ésima observación de la variable independiente del nivel 1

Z_{1j} = j – ésima observación de la variable independiente del nivel 2

β_k = Coeficiente de regresión en el nivel 1

ε_i = Error aleatorio en el nivel 1 ; $i = 1, \dots, n$.

σ_e^2 = Varianza del error aleatorio en el nivel 1.

1.3.3.2 *Modelo con intercepto aleatorio*

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij} ; \quad i = 1, \dots, n_j ; \quad j = 1, \dots, J \dots \dots \dots (I. 9)$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j} ; \quad j = 1, 2, \dots, J$$

$$\beta_{1j} = \gamma_{10} ; \quad j = 1, 2, \dots, J$$

Fuentes de variabilidad:

$$\varepsilon \sim N(0, \sigma_\varepsilon^2); \quad \mu \sim N(0, \sigma_\mu^2); \quad \text{Cov}(\varepsilon, \mu) = 0$$

Donde:

y_{ji} = i – ésima observación del nivel 1 de la j – ésima observación del nivel 2 de la variable dependiente

γ_{k0} = Coeficiente de regresión en el nivel 2

X_{ij} = i – ésima observación del nivel 1 de la j – ésima observación del nivel 2 de la variable independiente

β_k = Coeficiente de regresión en el nivel 1

ε_i = Error aleatorio en el nivel 1 ; $i = 1, \dots, n$.

σ_ε^2 = Varianza del error aleatorio en el nivel 1.

μ_{0j} = Error aleatorio en el nivel 2 ; $j = 1, \dots, J$.

σ_μ^2 = Varianza del error aleatorio en el nivel 2.

$\text{Cov}(\varepsilon, \mu)$ = Covarianza entre el error aleatorio del nivel 1 y 2.

Otra forma de expresar el modelo sería:

$$y_{ij} = (\gamma_{00} + \mu_{0j}) + \gamma_{10}X_{ij} + \varepsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \dots \dots \dots (I. 10)$$

En este modelo el intercepto varia pero la pendiente es constante.

1.3.3.3 Modelo con pendiente e intercepto aleatorios

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \dots \dots \dots (I. 11)$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j}; \quad j = 1, 2, \dots, J$$

$$\beta_{1j} = \gamma_{10} + \mu_{1j}; \quad j = 1, 2, \dots, J$$

Fuentes de variabilidad:

$$\varepsilon \sim N(0, \sigma_e^2)$$

$$\mu = \begin{pmatrix} \mu_{0j} \\ \mu_{1j} \end{pmatrix} \sim N\left(0, \Sigma_\mu = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right)$$

$$\text{Cov}(\varepsilon, \mu_k) = 0$$

Donde:

y_{ji} = i – ésima observación del nivel 1 de la j – ésima observación del nivel 2 de la variable dependiente

γ_{k0} = Coeficiente de regresión en el nivel 2

X_{ij} = i – ésima observación del nivel 1 de la j – ésima observación del nivel 2 de la variable independiente

β_k = Coeficiente de regresión en el nivel 1

ε_i = Error aleatorio en el nivel 1 ; $i = 1, \dots, n$.

σ_e^2 = Varianza del error aleatorio en el nivel 1.

μ_{kj} = Error aleatorio en el nivel 2 ; $j = 1, \dots, J$.

$$\Sigma_\mu = \begin{pmatrix} \sigma_{\mu_0}^2 = \tau_{00} & \sigma_{\mu_0, \mu_1} = \tau_{01} \\ \sigma_{\mu_0, \mu_1} = \tau_{01} & \sigma_{\mu_1}^2 = \tau_{11} \end{pmatrix}$$

= matriz de Varianza y Covarianza del error aleatorio del nivel 2.

$\text{Cov}(\varepsilon, \mu_k)$ = Covarianza entre el error aleatorio del nivel 1 y 2.

Otra forma de expresar el modelo sería:

$$y_{ij} = (\gamma_{00} + \mu_{0j}) + (\gamma_{10} + \mu_{1j})X_{ij} + \varepsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \dots (I.12)$$

1.3.4 Modelo de regresión simple con variables individuales (X) y contextuales (Z)

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \dots \dots \dots (I.13)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + \mu_{0j}; \quad j = 1, 2, \dots, J$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_{1j} + \mu_{1j}; \quad j = 1, 2, \dots, J$$

Fuentes de variabilidad:

$$\varepsilon \sim N(0, \sigma_e^2)$$

$$\mu = \begin{pmatrix} \mu_{0j} \\ \mu_{1j} \end{pmatrix} \sim N \left(0, \Sigma_\mu = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right)$$

$$\text{Cov}(\varepsilon, \mu_k) = 0$$

Donde:

y_{ji} = i – ésima observación del nivel 1 de la j – ésima observación del nivel 2 de la variable dependiente

γ_{kh} = Coeficiente de regresión en el nivel 2

X_{ij} = i – ésima observación del nivel 1 de la j – ésima observación del nivel 2 de la variable independiente

β_k = Coeficiente de regresión en el nivel 1

ε_i = Error aleatorio en el nivel 1 ; $i = 1, \dots, n$.

σ_e^2 = Varianza del error aleatorio en el nivel 1.

μ_{kj} = Error aleatorio en el nivel 2 ; $j = 1, \dots, J$.

$$\Sigma_\mu = \begin{pmatrix} \sigma_{\mu_0}^2 = \tau_{00} & \sigma_{\mu_0, \mu_1} = \tau_{01} \\ \sigma_{\mu_0, \mu_1} = \tau_{01} & \sigma_{\mu_1}^2 = \tau_{11} \end{pmatrix}$$

= matriz de Varianza y Covarianza del error aleatorio del nivel 2.

$\text{Cov}(\varepsilon, \mu)$ = Covarianza entre el error aleatorio del nivel 1 y 2.

Otra forma de expresar el modelo es:

$$y_{ij} = (\gamma_{00} + \gamma_{01}Z_{1j} + \mu_{0j}) + (\gamma_{10} + \gamma_{11}Z_{1j} + \mu_{1j})X_{ij} + \varepsilon_{ij}$$

$$y_{ij} = \gamma_{00} + \gamma_{01}Z_{1j} + \gamma_{10}X_{ij} + \gamma_{11}Z_{1j}X_{ij} + \mu_{0j} + \mu_{1j}X_{ij} + \varepsilon_{ij};$$

$$i = 1, \dots, n_j ; \quad j = 1, \dots, J \dots \dots \dots (I.14)$$

1.4 Estimación de parámetros en Modelos Multinivel

En un modelo de regresión jerárquico de dos niveles existen 3 tipos de parámetros que deben ser estimados a partir de los datos:

- Los parámetros de efectos fijos: $\gamma = (\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11})$
- Los coeficientes aleatorios de nivel 1: $\beta = (\beta_{0j}, \beta_{1j})$
- La varianza del error de nivel 1 : (σ_e^2) y la matriz de covarianzas de los errores de nivel 2 : $\Sigma_\mu = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}$

La estimación de cada uno de estos parámetros depende de los otros.

Todos los supuestos para el análisis de regresión lineal también son válidos para el análisis multinivel lineal (cuya función de enlace es la identidad). La variable dependiente se debe distribuir normalmente. El supuesto de independencia de las observaciones no se aplica en estos modelos, porque la razón para realizar un análisis multinivel es que las observaciones de los datos que van a ser analizados estén correlacionados.^[8]

1.5 Coeficiente de Correlación Intraclass (CCI)

El Coeficiente de Correlación Intraclass CCI, (The intraclass coefficient ICC), también es conocido como Coeficiente de Partición de la Varianza CPV, (Variance Partition Coefficient VPC). Este coeficiente es una proporción la cual es resultado

de un cociente de varianzas en el cual su denominador es la varianza total, se puede entender como el porcentaje de varianza que se desea explicar en determinado nivel respecto al porcentaje de varianza total. Este coeficiente se denota como " ρ " su valor se encuentra entre cero y uno.

A partir de ello vamos a encontrar en las diferentes bibliografías representar el " ρ " de la siguiente manera:

$$\rho = \frac{\sigma_{\mu}^2}{\sigma_{\mu}^2 + \sigma_{\varepsilon}^2} ; \quad 0 \leq \rho \leq 1 \dots \dots \dots (I. 15)$$

La fórmula I.15 se define en el caso de un modelo con sólo una intersección aleatoria, la variación en la intersección aleatoria en la unidad de nivel 2, y la variación residual a nivel 1. ^[9]

La correlación intraclase es el grado de semejanza entre las micro-unidades que pertenecen a la misma macro-unidad esta puede ser expresada por el coeficiente de correlación intraclase. Sin embargo pueden haber varias definiciones de este coeficiente, dependiendo del diseño muestral (en este caso un diseño de muestreo en dos etapas). ^[10]

Un modelo relevante aquí es el modelo de efectos aleatorios ANOVA en el cuál tenemos una variable respuesta y_{ij} , que viene a ser el valor observado en la macro unidad i dentro de la macro unidad j , este modelo se puede expresar como:

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

Donde:

- μ : *es la población media general.*
- τ_j : *es el efecto específico de la macro-unidad j .*
- ε_{ij} : *es el efecto residual de la micro-unidad i .*

De ahí que podemos obtener el coeficiente de correlación intraclase como:

$$CCI = \rho = \frac{\text{varianza poblacional entre las macro unidades}}{\text{varianza total}} = \frac{\sigma_{\tau}^2}{\sigma_{\tau}^2 + \sigma_{\varepsilon}^2} \dots (I. 16)$$

σ_{τ}^2 : Es la varianza entre los grupos
(es decir la varianza entre las macro-unidades).

σ_{ε}^2 : Es la varianza dentro de los grupos
(es decir la varianza de los sujetos dentro de la micro-unidad)

Por tanto la varianza total sería la suma de estas dos componentes $Var(Y_{ij}) = \sigma_{\tau}^2 + \sigma_{\varepsilon}^2$

Por ejemplo asumamos que tenemos un modelo de dos niveles:

$$CCI = \rho = \frac{\text{Variación de nivel 2}}{\text{Variación de nivel 1} + \text{Variación de nivel 2}} \dots \dots (I. 17)$$

En este caso el coeficiente de correlación intraclase representa el porcentaje de varianza explicada por el nivel 2. Recordemos el caso de un modelo con intercepto aleatorio, se mide la correlación residual entonces se entenderá variación del nivel 2 (varianza residual de segundo nivel) y variación del nivel 1 (varianza residual del primer nivel).

En un modelo de tres niveles (por ejemplo: estudiantes, aulas y escuelas) vamos a obtener dos de estos coeficientes de correlación, una correlación intraclase para el tercer nivel (que es la proporción de la varianza que se sitúa entre las escuelas) y otra correlación intraclase para el segundo nivel (que es la proporción de la varianza que se sitúa entre las aulas). ^[11]

Modelización multinivel permite estimar fácilmente coeficientes de correlación intraclase (ICC), que pueden alterar sustancialmente los términos de error en los

análisis y así crear falsos positivos. Una regla común es el uso de los modelos multinivel al ICC es mayor que 0,05.^[12]

1.6 Funciones de enlace para variables de respuesta binaria

La función de enlace proporciona la relación entre el predictor lineal (η) y la función de distribución. En la siguiente tabla se presenta la relación entre las distribuciones muestrales y las funciones de enlace.^[9]

Cuadro N° I.18. FUNCIONES DE ENLACE

<div>Enlace</div> <div>Distribución</div>	Log – Log Complementario	Identidad	Log	Logit	OLog – Log Complementario	OLogit	OProbit	Potencia	Probit
Bernoulli	✓			✓					✓
Binomial	✓			✓					✓
Gamma			✓					✓	
Gaussiana Inversa			✓					✓	
Multinomial				✓	✓	✓	✓		
Binomial Negativa			✓						
Normal		✓							
Poisson			✓						

En la presente investigación sólo interesan las funciones de enlace de respuesta binaria. Estas funciones del enlace transforman el valor observado en η y asegurar que la probabilidad predicha se encuentra dentro del intervalo (0,1).

Las funciones de enlace disponibles para los resultados binarios con distribución Bernoulli en MGLIM incluyen funciones logit, probit y log-log (Ver

el cuadro I.18). En el cuadro I.19 se aprecia para cada función de enlace sus correspondientes funciones de distribución acumulativa (CDF), las cuales son las funciones inversas de cada uno, y su media y la varianza.

Cuadro N° I.19. FUNCIONES DE ENLACE DE RESPUESTA BINARIA

Nombre del Enlace	Función del enlace ($F^{-1}(p) = \eta$) $0 < p < 1$	Función de distribución acumulativa ($F(\eta)$) $-\infty < \eta < \infty$	Media	Varianza
Logit (Logística)	$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$	$\frac{e^\eta}{1 + e^\eta}$	0	$\frac{\pi^2}{3}$
Probit	$\Phi^{-1}(p)$ Donde Φ^{-1} es la inversa de la distribución acumulada de la distribución normal estándar.	$\Phi(\eta)$	0	1
Log – Log Complementario	$\log(-\log(1 - p))$	$1 - e^{-e^\eta}$	-0.577	$\frac{\pi^2}{6}$

Estas funciones de enlace asignan la probabilidad de η dentro del intervalo abierto (0,1) de los números reales. La función de enlace logit y probit son simétricos. La función logit tiene una variación más grande. La función de enlace log-log complementario es asimétrica. Cuando p es extremadamente grande o pequeño(s), la relación lineal no se mantiene. Comprender la función de enlace es esencial a la hora de interpretar los resultados. ^[9]

Algunos de los Algoritmos empleados para la estimación de parámetros en los Modelos Multinivel están basados en el método de máxima verosimilitud o también en los métodos de Montecarlo, en estos casos es necesario los programas computacionales por su complejidad.

Algoritmos basados en el método de máxima verosimilitud.

- a. Máxima verosimilitud Completa (FML)
 - El método de mínimos cuadrados generalizados iterativo (IGLS).
 - El método Score de Fisher.
- b. Método de Máxima verosimilitud restringida (RML- MVR)
 - Mínimos cuadrados generalizados iterativos restringidos (RIGLS) (Goldstein (1989)).
 - Algoritmo EM (Dempster, Laird y Rubin (1977)) utilizado en el análisis de datos categóricos.

Algoritmos basados en métodos de montecarlo (cadenas de Markov)

Una alternativa a los métodos de máxima verosimilitud son los algoritmos basados en los métodos de Montecarlo, la importancia de este método radica en que puede generar aproximaciones bastante cercanas a las funciones de densidad de las variables estudiadas, por ende es posible realizar estimaciones puntuales como intervalares para los parámetros de los modelos. La utilización de este método en comparación de los de máxima verosimilitud en el análisis multinivel se justifica en que:

- Se pueden utilizar para obtener intervalos de confianza bastante exactos.
- Pueden ser utilizados para ajustar modelos en el contexto bayesiano.

En particular cuando tenemos una variable respuesta cualitativa de tipo dicotómica la cual se quiere predecir en función de variables explicativas cualitativas y/o cuantitativas los modelos logit y probit son los más utilizados como se ha visto anteriormente. (Ver cuadro I.19), a partir de ello es posible evaluar la influencia de una o más variables sobre un evento de interés a

través de las prueba de hipótesis e intervalos de confianza con la inferencia basada en el estadístico de wald o bien las estadísticas basadas en la razón de verosimilitud. Esto último ha llamado la atención a muchos investigadores el cual se han propuesto como objetivo evaluar y comparar el desempeño de estas inferencias a través de los métodos de Montecarlo.^[13]

Paralelamente de ser de interés la revisión de estos algoritmos informáticos se cuenta con información adicional de los métodos de simulación de Monte Carlo con cadenas de Markov y algunas variaciones como el algoritmo de Hastings–Metropolis, el muestreador de Gibbs.^[14]

1.7 Validación del Modelo

Para poder validar el modelo se realizará la evaluación de la significancia del modelo pero a partir de los coeficientes en conjunto, es decir determinar si las variables independientes son significativas o no, se plantea las siguientes hipótesis:

Hipótesis

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \dots = \beta_k = 0 \\ H_1 : \text{Por lo menos un } \beta_i \neq 0 \quad \forall i = 1, 2, \dots, k \end{cases}$$

Nivel de significación: α

Estadístico de Prueba:

El estadístico que se plantea para la evaluación de la significancia del modelo es la diferencia del valor de la desviación del modelo sólo con la constante y del modelo incluyendo las variables independientes, este

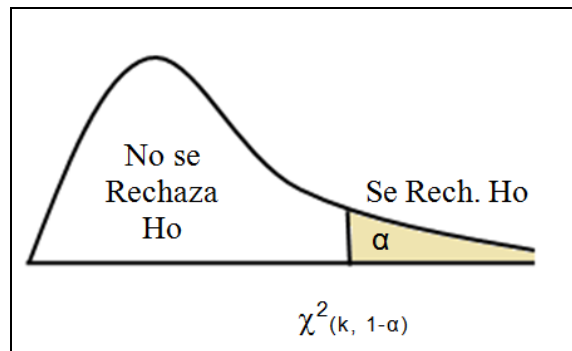
estadístico sigue una distribución Chi – Cuadrada con k grados de libertad, se tiene:

$$G = D(\text{modelo sin variables}) - D(\text{modelo con variables})$$

$$G = -2\ln\left(\frac{\text{verosimilitud del modelo sin variables}}{\text{verosimilitud del modelo con variables}}\right) \dots \dots \dots (I. 20)$$

Criterio de Decisión:

Figura N° I.21. Representación Gráfica de la zona de rechazo de la validación del modelo



Donde:

$\chi^2_{(k, 1-\alpha)}$ es un Valor crítico.

k = número de variables

α = nivel de significación

Si: $G > \chi^2_{(k, 1-\alpha)}$ entonces se rechaza H_0 , por lo tanto una o más de las variables Independientes consideradas en el Modelo son significativos.

1.7.1 Evaluación de la capacidad predictiva del modelo

Desvianza del Modelo (D):

Para evaluar la capacidad predictiva del modelo, es decir determinar si el modelo ajustado es el adecuado o no (bondad de ajuste), se plantea las siguientes hipótesis:

Hipótesis

$$\begin{cases} H_0 : \text{El modelo ajustado es significativo.} \\ H_1 : \text{El modelo ajustado no es significativo.} \end{cases}$$

Nivel de significación: α .

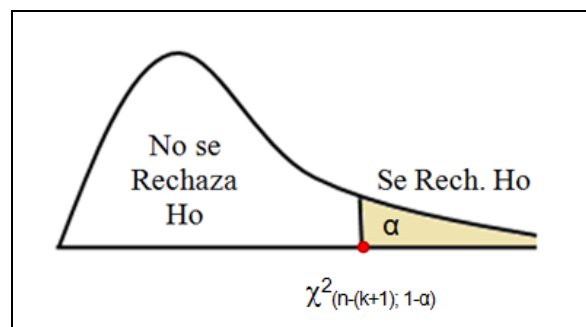
Estadístico de Prueba:

La desvianza es la medida del grado de diferencia entre las frecuencias predichas y las observadas del modelo, **el mejor modelo será aquel que tenga menor desvianza**^[15], este estadístico sigue una distribución Chi – Cuadrada con $n - (k+1)$ grados de libertad, se tiene:

$$D = -2\ln \left(\frac{\text{verosimilitud del modelo ajustado}}{\text{verosimilitud del modelo saturado}} \right) \dots \dots \dots (I. 22)$$

Criterio de Decisión:

Figura N° I.23. Representación Gráfica de la zona de rechazo de la desvianza del modelo



Dónde:

$\chi^2_{(n-(k+1); 1-\alpha)}$ = Valor crítico.

n = número de registros

k = número de variables

α = nivel de significación

Si $D < \chi^2_{(n-(k+1); 1-\alpha)}$ entonces No se rechaza la H_0 , por tanto el modelo ajustado es significativo, es decir no existe diferencia entre los valores observados y los valores estimados.

1.7.2 Prueba de hipótesis individual de las variables

Esta prueba individual de las variables se refiere a la significancia de los coeficientes del modelo. Es decir se está considerando una prueba individual de las variables para ver la importancia de la misma en el modelo.

HIPOTESIS

Se plantea las siguientes hipótesis:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

Con un **nivel de significación de α**

ESTADÍSTICOS DE PRUEBA:

$$Z_c = \frac{\hat{\beta}_j}{\widehat{EE}(\hat{\beta}_j)} \sim N(0,1) \dots \dots \dots (I.24)$$

ó

$$W_c = \frac{\hat{\beta}_j^2}{\widehat{Var}(\hat{\beta}_j)} \sim \chi_1^2 \dots \dots \dots (I.25)$$

CRITERIO DE DECISIÓN:

Se rechaza H_0 , si: $|Z_c| > Z_{1-\alpha/2}$ ó $W_c > \chi_{1,1-\alpha}^2$

El estadístico W_c es denominado Estadístico de Wald.

La fórmula I.25 es el estadístico de Wald, el cual es una aproximación de la fórmula I.24, que usa distribución normal estándar, esto será demostrado con el siguiente teorema:

Teorema: Sea Z una variable aleatoria con distribución normal estándar. La variable aleatoria $X = Z^2$ tiene distribución chi cuadrada con un grado de libertad.

Para poder comprobar la afirmación del teorema anterior procederemos a desarrollar de manera detallada su demostración.

Demostración:

$$X = Z^2$$

$$Z = \pm\sqrt{X}$$

Derivando:

$$\frac{dz}{dx} = \pm \frac{1}{2\sqrt{X}}$$

Entonces aplicando el método Jacobiano, tenemos:

$$f_X(x) = \left(f_Z(-\sqrt{X}) + f_Z(\sqrt{X}) \right) \times \left| \frac{dz}{dx} \right|$$

Recordemos que: $Z \sim N(0,1)$, entonces su función de densidad es:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Reemplazándolo en la ecuación anterior, se tiene:

$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \right) \times \left| \pm \frac{1}{2\sqrt{x}} \right|$$

$$f_X(x) = \left(\frac{2}{\sqrt{2\pi}} e^{-\frac{x}{2}} \right) \times \frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}$$

Como se sabe que: $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, luego:

$$f_X(x) = \frac{1}{\Gamma\left(\frac{1}{2}\right)\sqrt{2x}} e^{-\frac{x}{2}} = \frac{1}{\Gamma\left(\frac{1}{2}\right)\sqrt{2}} e^{-\frac{x}{2}} \times (x)^{-\frac{1}{2}}$$

$$f_X(x) = \frac{1}{\Gamma\left(\frac{1}{2}\right)\sqrt{2}} e^{-\frac{x}{2}} \times \left(2\frac{x}{2}\right)^{-\frac{1}{2}} = \frac{1}{\Gamma\left(\frac{1}{2}\right)\sqrt{2}\sqrt{2}} e^{-\frac{x}{2}} \times \left(\frac{x}{2}\right)^{-\frac{1}{2}}$$

Finalmente:

$$f_X(x) = \frac{1}{2 \times \Gamma\left(\frac{1}{2}\right)} e^{-\frac{x}{2}} \times \left(\frac{x}{2}\right)^{\frac{1}{2}-1}$$

Por lo tanto: $X \sim \chi^2_{1gl}$

Con esta demostración se concluye que al elevar al cuadrado el estadístico obtenido bajo la distribución normal estándar obtendremos el estadístico para la distribución de wald, la cual es utilizada también como alternativa para conocer la importancia de la variable en el modelo, es decir al utilizar el estadístico Z, lograremos llegar a la misma conclusión que se obtiene con un estadístico de Wald.

1.8 Medidas de Calidad para el Ajuste Global de los Modelos

Se presentan los estadísticos de ajuste global cuando se desea cuantificar en qué medida un modelo representa la variabilidad de los datos (variaciones observadas en los datos), sin embargo estos indicadores no son directamente interpretables, su importancia radica en la comparación de los modelos. Los estadísticos de ajuste global son:

1.8.1 Criterio de Información de Akaike AIC, por Akaike en 1974

$$\text{AIC} = -2LL + 2d \dots \dots \dots (I. 26)$$

Se utiliza en la regresión aunque originalmente fue propuesto para los modelos de series de tiempo también es denotado como **-2lnL + 2r**, (donde se observa según el caso anterior que $d=r$, este representa el número de parámetros estimados en el modelo. El modelo con AIC mínimo, en un conjunto de modelos anidados, será el más parsimonioso de acuerdo con este criterio. ^[16]

1.8.2 Criterio de Información de Akaike corregido AICC, por Hurvich y Tsai, 1989. ^[16]

$$\text{AICC} = -2LL + \frac{2dn}{n - d - 1} \dots \dots \dots (I. 27)$$

1.8.3 Criterio de Información de Akaike consistente CAIC, por Bozdogan, 1987. ^[16]

$$\text{CAIC} = -2LL + d[\log(n) + 1] \dots \dots \dots (I. 28)$$

1.8.4 Criterio de Información Bayesiano *BIC*, por Schwarz, en 1978

$$\mathbf{BIC} = -2LL + d\log(n) \dots \dots \dots (I.29)$$

A menor valor de este criterio indicaría el modelo más parsimonioso entre los modelos comparados. Por lo general n indica el número de unidades en el nivel más alto de la jerarquía, sin embargo se debe tener cuidado ya que finalmente se debe tener en cuenta la estimación que se está realizando o el software con el que se está trabajando. ^[16]

Al comparar modelos alternativos esperamos que estos estadísticos mencionados anteriormente sean lo menor posible. Detallamos sus componentes de la siguiente manera:

LL : Logaritmo de la verosimilitud

$$= \ln \left(\frac{\text{verosimilitud del modelo ajustado}}{\text{verosimilitud del modelo saturado}} \right)$$

-2LL : También denotado como $D = -2\ln L$

Desvianza (Tomando en cuenta dos modelos alternativos, la diferencia entre la desvianza correspondientes a estos dos modelos mencionados, tiene una distribución X^2 donde g.l. es el número de parámetros estimados (Ver fórmula I.22).

Para comprobar los estadísticos, es importante tomar en cuenta el método de estimación que se está realizando. Por ejemplo: los

métodos de estimación puede ser el de Máxima Verosimilitud (M.V.) o bien el método de Máxima Verosimilitud Restringida (M.V.R.).^[16]

Si: M.V. entonces:

$d = n^{\circ}$ de parámetros de efectos fijos
+ n° de parámetros de efectos aleatorios
 $n = n^{\circ}$ total de casos.

Si: M.V.R. entonces:

$d = n^{\circ}$ de parámetros asociados al de efectos aleatorios
 $n = n^{\circ}$ total de casos – n° parámetros de efectos fijos

1.9 Ventajas

- Se diferencian dos tipos de variables: las "entre-grupo" o de nivel macro, cuyos valores varían de grupo a grupo (siendo el mismo para todas las unidades de un mismo grupo), y las "intra-grupo" o de nivel micro (cuyos valores varían entre las unidades del mismo grupo). Estos modelos jerárquicos permiten examinar simultáneamente el efecto de variables de nivel macro y micro sobre respuestas a nivel micro, considerando la falta de independencia de las respuestas dentro de los grupos. Esto es posible mediante la descomposición de la variabilidad total en una variación intra-grupo y otra variación entre-grupos.^[17]
- La incorporación de los denominados efectos aleatorios en el Modelo Multinivel, los cuales al no ser fijos sino variables – estocásticas – con una distribución de probabilidad asociada, permite especificar indirectamente la forma de asociación más pertinente a las relaciones que se buscan.^{[18] y [19]}

- Una ventaja del enfoque de modelización multinivel es que se puede tratar con los datos en los que los tiempos de las mediciones varían de un sujeto a otro.^[20]

1.10 Desventajas

- Tiene una mayor dificultad y complejidad en el cálculo del tamaño muestral la que depende tanto de la cantidad de grupos como unidades dentro de cada grupo.^[17]
- Presenta mayor complejidad tanto en el marco teórico como el modelo propuesto para analizar los datos.^[21]
- La interpretación de los resultados de un modelo multinivel no siempre es evidente, especialmente cuando las estructuras complejas de la variación son utilizadas.^[22]

CAPÍTULO II:

MODELO LOGIT

2.1 Introducción

El modelo logit permite obtener estimaciones de la probabilidad de un suceso e identificar los factores de riesgo que determinan dichas probabilidades.

Este modelo es empleado cuando se desea comparar un grupo de variables independientes con una variable dependiente no métrica (dicotómica), por lo que su posible respuesta es 0 (Fracaso, Ausencia de ...) o 1 (Éxito, Presencia de ...), por este motivo no se puede usar un modelo lineal.

La modelización Logit es similar a la regresión tradicional salvo que utiliza como función de estimación la función logística en vez de la lineal. Con la modelización Logit, el resultado del modelo es la estimación de la probabilidad de que un nuevo individuo pertenezca a un grupo o a otro, mientras que por otro lado, al tratarse de un análisis de regresión, también permite identificar las variables más importantes que explican las diferencias entre grupos.^[23]

2.2 Definición

Sea la variable respuesta Y , la cual sólo puede tomar los valores $Y=1$ (presencia de la característica de interés) con probabilidad de ocurrencia igual a π y $Y=0$ (ausencia de la característica de interés) con probabilidad $1-\pi$. Además sea la covariable (variable predictiva o variable independiente) X , la cual puede ser categórica o continua.

Si la variable Y es el resultado de un experimento de bernoulli, esto es, las observaciones son independientes, entonces la variable aleatoria tiene distribución de Bernoulli con:

$$E(Y / X = x) = \pi \text{ (Esperanza condicional de } Y \text{ dado } X=x), \text{ y}$$

$$V(Y / X = x) = \pi(1 - \pi) \text{ (Varianza condicional de } Y \text{ dado } X=x).$$

Por lo tanto, **la probabilidad de que $Y = 1$** es igual a la $E(Y / X = x) = \pi$ y puede ser calculada a partir de una distribución de probabilidad que tiene la forma de la curva sigmoidea, en particular esta curva puede ser la logística,

$$E(Y / X = x) = \pi = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \dots \dots \dots (II.1)$$

Donde $\eta = \beta_0 + \beta_1 X$ es el predictor lineal y la función de enlace canónico es:

$$\theta = \ln\left(\frac{E(Y)}{1 - E(Y)}\right) = \ln\left(\frac{\pi}{1 - \pi}\right) \dots \dots \dots (II.2)$$

Su representación como un modelo lineal generalizado, será:

$$\text{logit}(\pi) = \ln\left(\frac{E(Y/X = x)}{1 - E(Y/X = x)}\right) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 X \dots \dots \dots (II.3)$$

Este modelo es útil en situaciones prácticas de investigación en que la variable respuesta puede tomar solo dos valores, por ejemplo: desaprobado o aprobado; e interesa conocer la probabilidad de que un alumno este desaprobado en función de su perfil de variables predictivas o factores de riesgo.

La utilidad del modelo se basa en que muchas veces, el perfil de variables predictivas puede estar formado por características cualitativas y cuantitativas; y se pretende hacer participar a todas en una sola ecuación conjunta que explique como la probabilidad de alcanzar una respuesta depende de todas y cada una de las variables predictivas.

2.3 Función de Enlace LOGIT

La función de enlace Logit será:

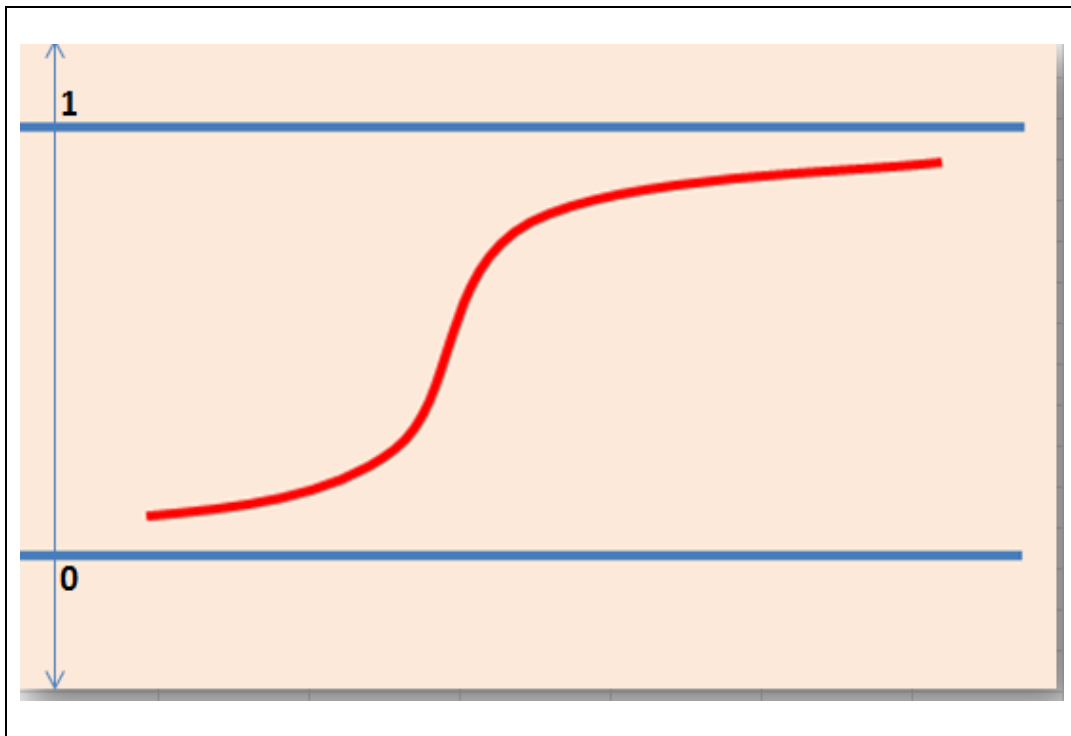
$$\pi_{ij} = \text{Prob}(Y_{ij}) = \frac{1}{1 + e^{-\eta_{ij}}} \dots \dots \dots (II.4)$$

Donde:

$$\eta_{ij} = \text{Logit}(\pi_{ij}) = \ln\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) \dots \dots \dots (II.5)$$

Es el predictor lineal, el cual representa el logaritmo del Odds del suceso o evento de interés.

Figura N° 2.6. Representación Gráfica de una función logística



Como se puede observar esta función tiene una forma sigmoidea y está acotada entre cero y uno.

Con esta función de enlace, la probabilidad $\text{Prob}(Y_{ij} = 1|\beta)$ se encontrara en el intervalo de 0 a 1.

Según Rasbash et al (2004), en una distribución logística los residuos del nivel 1 poseen una media igual a cero y una variancia igual a $\frac{\pi^2}{3}$.^[24] (Ver cuadro I.19)

Y el modelo podrá ser expresado en dos niveles:

2.3.1 **MODELO NULO**

Nivel 1:

$$\eta_{ij} = \text{Logit}(\pi_{ij}) = \beta_{0j} + \varepsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \dots (II.7)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00}$$

Donde: $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma_e^2 = \frac{\pi^2}{3}$

2.3.2 **MODELO CON UN CRITERIO**

Nivel 1:

$$\eta_{ij} = \text{Logit}(\pi_{ij}) = \beta_{0j} + \varepsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \dots (II.8)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Donde: $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma_e^2 = \frac{\pi^2}{3}$ $\mu \sim N(0, \sigma_\mu^2)$

2.3.3 MODELO CON INTERCEPTO ALEATORIO Y VARIABLES CONTEXTUALES

Nivel 1:

$$\eta_{ij} = \text{Logit}(\pi_{ij}) = \beta_{0j} + \sum \beta_{ih} X_{hj} + \varepsilon_{ij};$$

$$i = 1, \dots, n_j; \quad j = 1, \dots, J \dots \dots (II.9)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \sum \alpha_{0k} Z_{kj} + \mu_{0j}$$

$$\beta_{ih} = \gamma_{i0}$$

Donde: $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma_e^2 = \frac{\pi^2}{3}$ $\mu \sim N(0, \sigma_\mu^2)$

$$\text{Cov}(\varepsilon, \mu) = 0$$

2.3.4 MODELO CON INTERCEPTO Y PENDIENTES ALEATORIAS

Nivel 1:

$$\eta_{ij} = \text{Logit}(\pi_{ij}) = \beta_{0j} + \sum \beta_{ih} X_{hj} + \varepsilon_{ij};$$

$$i = 1, \dots, n_j; \quad j = 1, \dots, J \dots \dots (II.10)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \sum \alpha_{0k} Z_{kj} + \mu_{0j}$$

$$\beta_{ih} = \gamma_{i0} + \sum \alpha_{ik} Z_{kj} + \mu_{ij}$$

Donde: $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma_e^2 = \frac{\pi^2}{3}$ $\mu \sim N(0, \Sigma_\mu)$

$$\Sigma_{\mu} = \begin{pmatrix} \sigma_{\mu_0}^2 = \tau_{00} & \sigma_{\mu_0, \mu_1} = \tau_{01} \\ \sigma_{\mu_0, \mu_1} = \tau_{01} & \sigma_{\mu_1}^2 = \tau_{11} \end{pmatrix}$$

$$\text{Cov}(\varepsilon, \mu_k) = 0$$

2.3.5 Coeficiente de Correlación Intraclass

El coeficiente de correlación intraclass conocido como ρ (ver fórmula I.15 y cuadro I.19) para el modelo Logit es:

$$\rho = \frac{\sigma_{\mu_0}^2}{\sigma_{\mu_0}^2 + \frac{\pi^2}{3}} \dots \dots \dots (II. 11)$$

2.4 Ventajas del modelo Logit

- Al ser la variable dependiente dicotómica no necesita que se cumple el supuesto de normalidad.^[25]
- En este modelo los coeficientes por sí solos no tienen una interpretación sencilla, no obstante, al ser exponenciados y luego restados en uno, su interpretación se vuelve más comprensible.^[25]
- En un modelo multinivel con función de enlace logit se puede obtener probabilidades, entonces es posible clasificar a un individuo a la pertenencia de un grupo u otro.^[26]

2.5 Desventajas del modelo Logit

- El tamaño de muestra debe ser grande, ya que tiene como método de estimación la máxima verosimilitud.^{[27] y [28]}

CAPÍTULO III:

MODELO PROBIT

3.1 Introducción

Fue incorporado como una alternativa log-lineal para manejar conjuntos de datos con variable respuesta categórica binaria.

El análisis Probit es usado para analizar muchas clases de experimentos tipo dosis-respuesta (binaria) en una variedad de campos. Por ejemplo, en marketing alguien puede estar interesado en modelar preferencias de clientes por determinados productos (Compra/no compra) a partir de la aplicación de comerciales televisivos cada cierto número de minutos (dosis). En ensayos clínicos donde puede interesar el alivio de una dolencia (Si/no) como resultado de suministrar diferentes dosis de un medicamento.

Este análisis es comúnmente aplicado en la toxicología, para determinar la toxicidad relativa de productos químicos a organismos vivos. Esto se realiza observando la respuesta de un organismo a varias concentraciones de cada uno de los productos químicos en cuestión y luego comparando las concentraciones en las cuales se encuentra una respuesta (Por ejemplo, la muerte). Por lo tanto la respuesta es siempre binomial y la relación entre la dosis y la respuesta es de tipo sigmoidea. La transformación Probit actúa como una transformación de sigmoideo a lineal para poder ajustar el modelo de regresión.

3.2 Definición

A diferencia de la regresión logit que utiliza una función de enlace del tipo, logaritmo natural de los Odds Ratio, la función de enlace Probit es la inversa de una distribución normal estándar acumulada ($N(0,1)$).

La función de distribución acumulada de la normal estándar (Φ) se expresa:

$$g = \Phi(\eta) = P(Z \leq \eta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\eta} e^{-\frac{t^2}{2}} dt \dots \dots \dots (III.1)$$

Donde:

$$\eta = X^T \beta$$

Sea la variable respuesta Y , la cual sólo puede tomar los valores $Y=1$ (presencia de la característica de interés) con probabilidad de ocurrencia igual a π . Luego, el modelo de regresión Probit se expresa como:

$$\pi = E(Y/x) = \Phi(\eta) = \Phi(x^T \beta) \dots \dots \dots (III.2)$$

Para linealizar la relación aplicamos la transformación inversa:

$$\Phi^{-1}(\pi) = \Phi^{-1}\left(E\left(\frac{Y}{x}\right)\right) = \eta = x^T \beta \dots \dots \dots (III.3)$$

Donde $\eta = x^T \beta$ es el predictor lineal y la función de enlace canónico es:

$$\theta = \Phi^{-1}(\pi) = \Phi^{-1}(E(Y)) \dots \dots \dots (III.4)$$

Su representación como un modelo lineal generalizado, será:

$$probit(\pi) = \Phi^{-1}(\pi) = \Phi^{-1}(E(Y/x)) = X^T \beta \dots \dots \dots (III.5)$$

El modelo Probit permite llevar a cabo estudios en donde se cumple la existencia de una variable latente en donde se observa evidencia dicotómica.^[29]

Un modelo PROBIT de efectos aleatorios es desarrollado para los casos en donde la salida o resultado de interés es una serie correlacionada de la repuesta binaria.

GIBBONS Y BOCK (1997) han presentado un modelo PROBIT de efectos aleatorios para estimar la tendencia en una variable binaria de medidas repetidas en los mismos sujetos.

3.3 Función de Enlace PROBIT

La función de enlace probit será:

$$Prob(Y_{ij} = 1) = \Phi^{-1}(\eta_{ij}) \dots \dots \dots (III.6)$$

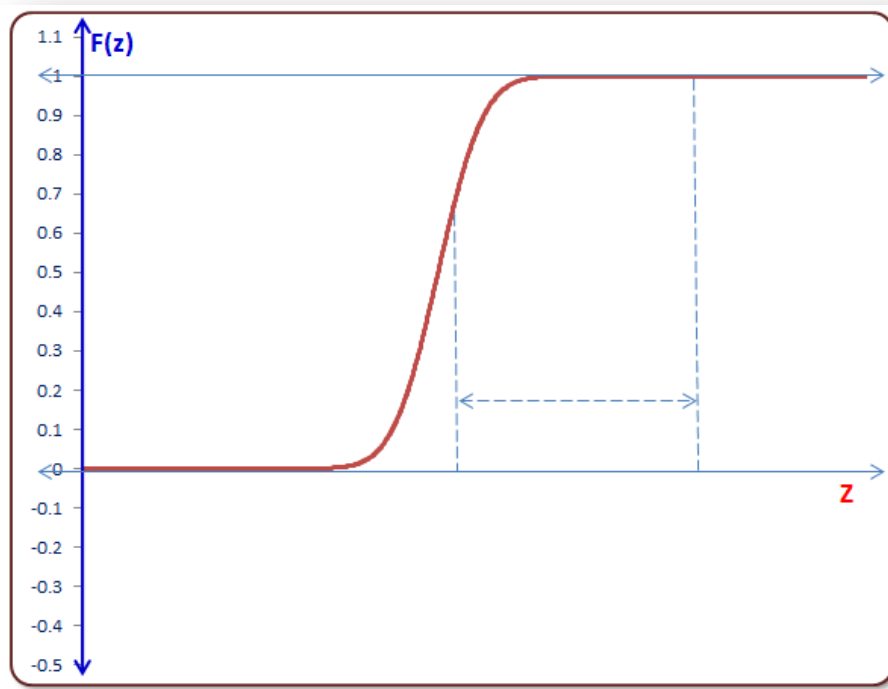
Donde:

$$\eta_{ij} = Probit(\pi_{ij}) = \Phi^{-1}(\pi_{ij}) \dots \dots \dots (III.7)$$

Es el predictor lineal, y Φ es la Distribución Acumulada Normal (0,1).

Con esta función de enlace, la probabilidad $Prob(Y_{ij} = 1|\beta)$ se encontrara en el intervalo de 0 a 1.

Figura N° 3.1. Representación Gráfica de una Distribución Normal Estándar Acumulada



Tanto la distribución normal estándar acumulada usada en Probit como la curva logística utilizada en la regresión logit tienen forma de S alargada, aunque la normal acumulada tiene una pendiente ligeramente mayor (pero en general las diferencias son muy pequeñas).

En una distribución probit los residuos del nivel 1 poseen distribución normal con una media igual a cero y una variancia igual a 1. (Ver cuadro I.19)

Y el modelo podrá ser expresado en dos niveles:

3.3.1 MODELO NULO

Nivel 1:

$$\eta_{ij} = Probit(\pi_{ij}) = \beta_{0j} + \varepsilon_{ij} ; \quad i = 1, \dots, n_j ; \quad j = 1, \dots, J \dots (III.8)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00}$$

Donde: $\varepsilon \sim N(0, \sigma_e^2 = 1)$

3.3.2 MODELO CON UN CRITERIO DE CLASIFICACIÓN

Nivel 1:

$$\eta_{ij} = Probit(\pi_{ij}) = \beta_{0j} + \varepsilon_{ij} ; \quad i = 1, \dots, n_j ; \quad j = 1, \dots, J \dots (III.9)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Donde: $\varepsilon \sim N(0, \sigma_e^2 = 1)$ $\mu \sim N(0, \sigma_\mu^2)$ $Cov(\varepsilon, \mu) = 0$

3.3.3 MODELO INTERCEPTO ALEATORIO Y VARIABLES CONTEXTUALES

Nivel 1:

$$\eta_{ij} = Probit(\pi_{ij}) = \beta_{0j} + \sum \beta_{ih} X_{hj} + \varepsilon_{ij} ;$$

$$i = 1, \dots, n_j ; \quad j = 1, \dots, J \dots \dots (III. 10)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \sum \alpha_{0k} Z_{kj} + \mu_{0j}$$

$$\beta_{ih} = \gamma_{i0}$$

$$\text{Donde:} \quad \varepsilon \sim N(0, \sigma_e^2 = 1) \quad \mu \sim N(0, \sigma_\mu^2) \quad \text{Cov}(\varepsilon, \mu) = 0$$

3.3.4 MODELO CON INTERCEPTO Y PENDIENTES ALEATORIAS

Nivel 1:

$$\eta_{ij} = \text{Probit}(\pi_{ij}) = \beta_{0j} + \sum \beta_{ih} X_{hj} + \varepsilon_{ij} ;$$

$$i = 1, \dots, n_j ; \quad j = 1, \dots, J \dots \dots (III. 11)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \sum \alpha_{0k} Z_{kj} + \mu_{0j}$$

$$\beta_{ih} = \gamma_{i0} + \sum \alpha_{ik} Z_{kj} + \mu_{ij}$$

$$\text{Donde:} \quad \varepsilon \sim N(0, \sigma_e^2 = 1) \quad \mu \sim N(0, \Sigma_\mu)$$

$$\Sigma_\mu = \begin{pmatrix} \sigma_{\mu_0}^2 = \tau_{00} & \sigma_{\mu_0, \mu_1} = \tau_{01} \\ \sigma_{\mu_0, \mu_1} = \tau_{01} & \sigma_{\mu_1}^2 = \tau_{11} \end{pmatrix}$$

$$\text{Cov}(\varepsilon, \mu_k) = 0$$

3.3.5 Coeficiente de Correlación Intraclase

El coeficiente de correlación intraclase (ρ) para el modelo Probit es:

$$\rho = \frac{\sigma_{\mu 0}^2}{\sigma_{\mu 0}^2 + 1} \dots \dots \dots (III. 12)$$

La representación para la segunda componente de la varianza dentro de la unidad del primer nivel se da en base a su respectiva función de enlace. (Ver formula I.15 y cuadro I.19).

3.4 Ventajas del modelo Probit

- Se pueden obtener estimaciones de probabilidad para la ocurrencia de un suceso.^[9]
- Construye una variable latente continua en la cual el investigador puede estar más interesado que en sus probabilidades.^[30]

3.5 Desventajas del modelo Probit:

- El tamaño de muestra debe ser grande, ya que tiene como método de estimación la máxima verosimilitud.^{[27] y [28]}
- Los coeficientes no tienen una interpretación directa.^[31]

*Las ideas son la fuente de
inspiración para nuevos
proyectos. A su vez los proyectos
se mantienen con nuevos
proyectos. V.H.M.S*

CAPITULO IV

En el presente capitulo se presenta una aplicación del análisis multinivel con la base de datos de PISA (PROGRAMA PARA LA EVALUACIÓN INTERNACIONAL DE ALUMNOS) 2009, no sin antes presentar el contexto del cual proviene estos datos, entre otros puntos muy importantes como sus variables de estudio, gestores y/o responsables, **la siguiente información es un extracto del Volumen I de PISA 2009.**

“INTRODUCCIÓN PISA 2009, es una iniciativa fruto de la colaboración entre los países participantes para unir el conocimiento científico bajo la dirección conjunta de sus gobiernos sobre la base de unos intereses comunes que pretenden desarrollar mejores políticas educativas.

Con los objetivos de la OCDE (ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICO) como contexto, la junta de gobierno de PISA, en la que todos los países están representados determina las políticas educativas prioritarias de PISA y supervisa que estas se tengan en cuenta durante la puesta en práctica del programa. Esto supone establecer las prioridades para el desarrollo de indicadores, elaborar los instrumentos de evaluación e informar de los resultados.

Además, expertos en materia de educación de los países participantes con una sólida experiencia a nivel internacional se ocupan, mediante grupos de trabajo, fijar los objetivos educativos. Al tomar parte en estos grupos de trabajo, los países se aseguran que los instrumentos son válidos a nivel internacional, se tienen en cuenta los contextos culturales y educativos de los países de la OCDE, los materiales de evaluación tienen una gran capacidad de medición y los instrumentos se centran en la autenticidad y la validez educativas.

Por medio de los coordinadores nacionales del proyecto, los países participantes ponen en marcha PISA a nivel nacional siguiendo los procedimientos administrativos acordados. Estos coordinadores nacionales desempeñan un papel fundamental, pues certifican que la puesta en marcha de la evaluación se está llevando a cabo de forma correcta y, posteriormente, verifican y evalúan los resultados, análisis, informes y publicaciones del estudio.

El diseño y la puesta en práctica de las evaluaciones, dentro del marco establecido por la Junta de Gobierno de PISA, es responsabilidad de colaboradores externos. El desarrollo de cuestionarios de PISA 2009 lo ha llevado a cabo un consorcio liderado por el Australian Council for Educational Research (ACER). Otros miembros de este consorcio son cApStAn Linguistic Quality Control en Bélgica, el Deutsches Institut für Internationale Pädagogische Forschung (DIPF) en Alemania, el National Institute for Educational Policy Research en Japón (NIER), la Unité d'analyse des systèmes et des pratiques d'enseignement (aSPe) en Bélgica y la empresa WESTAT en Estados Unidos.

El Secretariado de la OCDE es responsable de la gestión global del programa, controla el seguimiento diario de su puesta en práctica, ejerce como secretaría para la Junta de Gobierno de PISA, fomenta el consenso entre países y actúa de interlocutor entre la Junta de Gobierno de PISA y el consorcio internacional encargado de la aplicación de las actividades. Además, el Secretariado de la

OCDE elabora los indicadores y los análisis de los datos y prepara los informes y las publicaciones internacionales en cooperación con el Consorcio PISA y los países miembros, tanto en el ámbito de las políticas educativas (Junta de Gobierno de PISA) como en el de su implantación (coordinadores nacionales del proyecto)”. La lista de los miembros de los diversos organismos de PISA y los expertos asesores independientes que han colaborado con PISA, en los diferentes países incluyendo en el Perú.^[32]

Es importante mencionar a los países que son miembros de la OCDE: Alemania, Australia, Austria, Bélgica, Canadá, Chile, Corea, Dinamarca, Eslovenia, España, Estados Unidos, Estonia, Finlandia, Francia, Grecia, Hungría, Irlanda, Islandia, Israel, Italia, Japón, Luxemburgo, México, Noruega, Nueva Zelanda, Países Bajo, Polonia, Portugal, Reino Unido, República Checa, República Eslovaca, Suecia, Suiza y Turquía.

Como podemos notar el Perú no es miembro de la OCDE, sin embargo si ha participado en el año 2009, estando ausente en el año 2003 y 2006.

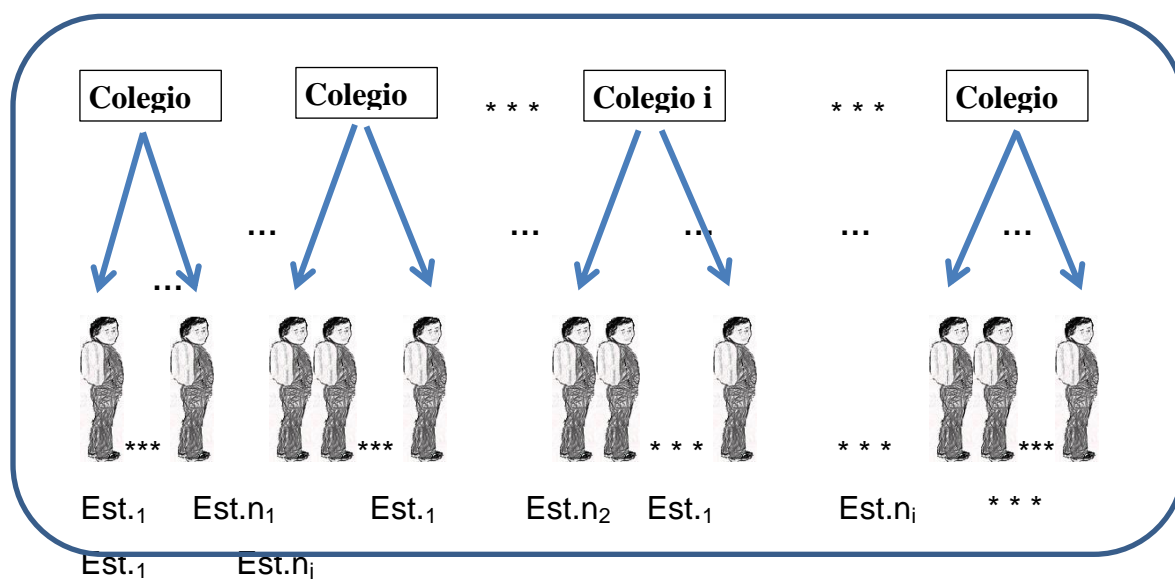
En el año 2009 PISA, evaluó principalmente la competencia lectora, pero sin dejar de lado las áreas de matemáticas y ciencias. Este proyecto, está dirigido a los estudiantes de 15 años de edad, examinando el grado en que los estudiantes han adquirido parte de los conocimientos y a su vez las capacidades que necesitan para desenvolverse en la sociedad, en particular en competencia lectora la preocupación es como analizan un texto escrito, la diversidad de enfoques que aportan a los textos, la capacidad de extraer la información de manera práctica concreta, como leer para aprender y comprender otras formas de hacer, pensar y ser. Con esta finalidad obtienen una serie de variables y a su vez construyen indicadores para ello utilizan tres cuestionarios dirigidos a los colegios, a los padres y a los estudiantes.

En el año mencionado, participaron miles de colegios en todo el mundo, estos fueron representados por una cantidad alrededor de medio millón de estudiantes (515 958). Y los puntajes promedios obtenidos en las diferentes competencias por los países miembros de la OCDE en el año 2009 fueron las siguientes:

Comprensión Lectora : 493 puntos
 Competencia Matemática : 496 puntos
 Competencia Científica : 501 puntos

En la presente tesis nos concentraremos en los colegios que participaron en el Perú por lo tanto los estudiantes que fueron evaluados en ellos. Entonces los estudiantes son nuestras unidades a nivel micro, y los colegios son nuestras unidades a nivel macro. Gráficamente tendríamos una estructura jerárquica de la siguiente forma:

GRÁFICO ESTRUCTURAL DE DOS NIVELES



En el Perú fueron seleccionados 240 colegios, que hacen un total de 5985 estudiantes. Observemos que no se tomaron todos los colegios del Perú para ello

se tomó una muestras de colegios al azar, como si tomáramos una muestra de tratamientos al azar de un gran conjunto de tratamientos recordemos que en el ANOVA a estos diseños se le denominan de efectos aleatorios.

4.1 VARIABLES E INDICADORES

En el siguiente cuadro se presentan algunas de las variables e indicadores utilizados en PISA 2009, para el nivel micro (NIVEL 1) y el nivel macro (NIVEL 2).

VARIABLES NIVEL 1

NOMBRE	ESPECIFICACIÓN / TRADUCCIÓN	TIPO DE VARIABLE
SCHOOLID	School ID 5-digit / Identificación del colegio	Cualitativa Nominal
StlDStd	Student ID 5-digit /Identificación del estudiante	Cualitativa Nominal
ST01Q01	Grade /Grado	Cualitativa Ordinal
ST04Q01	Sex /Sexo del estudiante	Cualitativa Nominal
METASUM	Meta cognition: Summarizing / Meta cognición: Resumen	Cuantitativa continua
UNDREM	Meta cognition: Understanding and Remembering / Meta cognición: Comprender y recordar	Cuantitativa continua
ESCS	Index of economic, social and cultural status (WLE) / Índice de estatus económico, social y cultural.	Cuantitativa continua
HISEI	Highest parental occupational status / Máximo nivel ocupacional de los padres	Cuantitativa discreta
PARED	Highest parental education in years / Máximo nivel educativo de los padres	Cuantitativa discreta
HOMEPOS	Home Possessions /Posesiones en el Hogar	Cuantitativa continua
WEALTH	Wealth/ Riqueza	Cuantitativa continua
HEDRES	Home educational resources/ Recursos educativos en el hogar	Cuantitativa continua
CULTPOSS	Cultural Possessions / Posesiones culturales	Cuantitativa continua
ONLNREAD	Online Reading / Lectura digital	Cuantitativa continua
JOYREAD	Joy/Like Reading / Leer por placer	Cuantitativa continua
LMINS	Learning time / Tiempo de aprendizaje	Cuantitativa continua
DISCLIMA	Disciplinary Climate/ ambiente disciplinario	Cuantitativa continua
PMat5	Promedio de Matemáticas	Cuantitativa continua
PCL5	Promedio de comprensión lectora	Cuantitativa continua
PCien5	Promedio de Ciencias	Cuantitativa continua
XW_FSTUWT	FINAL STUDENT WEIGHT /(Peso Final de Est-Col)	Cuantitativa continua

VARIABLES NIVEL 2

NOMBRE	ESPECIFICACIÓN / TRADUCCIÓN	TIPO DE VARIABLE
SCHOOLID	School ID 5-digit / Identificación del colegio	Cualitativa Nominal
SC02Q01	Public or private / Público o privado	Cualitativa Nominal
SC04Q01	School Community / comunidad en la que se encuentra ubicado el colegio.	Cualitativa Nominal
STRATUM	Original Stratum / El estrato original	Cualitativa Nominal
ZONA	Zona / La zona en la que está ubicada el colegio	Cualitativa Nominal
TC	Tipo de Colegio	Cualitativa Nominal
SC06Q01	Number of boys / número de alumnos	Cuantitativa discreta
SC06Q02	Number of girls / número de alumnas	Cuantitativa discreta
SC26Q11	Curriculum Responsibility/ Responsabilidad del currículo.	Cualitativa nominal
PROPQUAL	Proportion of qualified teachers/ Proporción de profesores calificados.	Cuantitativa continua
PROPCERT	Proportion of certified teachers/ Proporción de profesores certificados	Cuantitativa continua
SCMATEDU	Quality of the schools educational resources/ Recursos educativos de calidad en el colegio.	Cuantitativa continua
COMPWEB	Proportion of computers connected to the internet/ Proporción de computadoras conectadas a internet	Cuantitativa continua
ZW_FSCHWT	Final school weight / Peso Final del colegio	Cuantitativa continua

SCHOOLID: Código del colegio, en la base de datos solo se cuenta con una etiqueta de numeración para poder diferenciar y/o identificar al colegio, esta solo nos sirve de guía para saber cuántos colegios fueron seleccionados, o bien cuantos estudiantes participaron en el colegio. Este identificador no es un código modular ni nombre alguno del mismo.

STLDSTD: Código del estudiante, de la misma manera que el caso anterior es una etiqueta de numeración solo con el objetivo de diferenciar a los estudiantes, no es ningún código modular ni nombre del estudiante.

ST01Q01: Grado, aquí se muestra el grado que está cursando el estudiante, recordemos que si bien es cierto que los estudiantes tienen la misma edad no necesariamente se encuentran en el mismo grado.

ST04Q01: Sexo del estudiante. Femenino (Female) y Masculino (Male).

METACOGNICIÓN

El termino Meta cognición nace en la psicología especialmente en neuropsicología, también toma gran importancia en el campo de la educación, actualmente muchos especialistas en estas áreas mencionan inclusive a “teorías de la meta cognición”. En las diferentes bases de datos que tienen indexadas a diferentes revistas académicas de artículos científicos, se cuentan con muchas investigaciones del término Metacognición, es necesario la revisión de algunos artículos con la intención de tener mayor claridad respecto a esta variable que es parte de la base de datos de este capítulo de resultados.

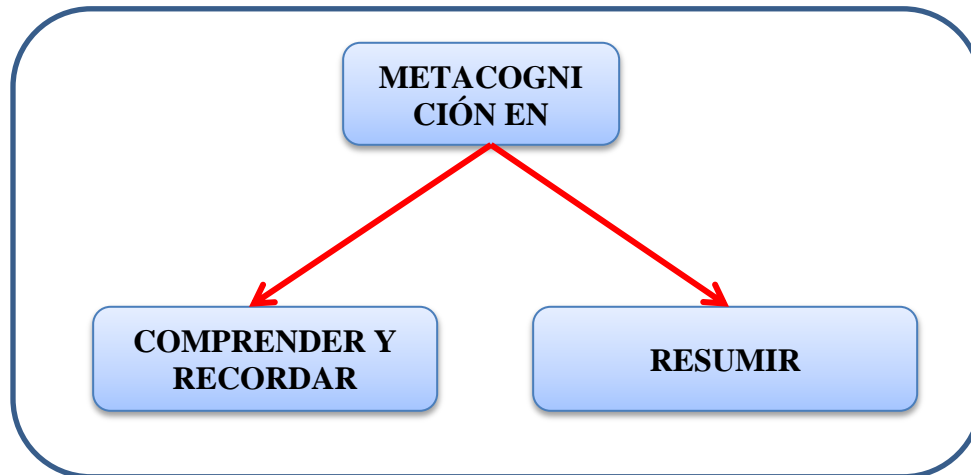
OCHOA Solanlly, ARAGÓN Lucero (2007); presentaron la investigación, “Funcionamiento Meta cognitivo de estudiantes universitarios durante la escritura de reseñas analíticas” del cual participaron 33 estudiantes de psicología, quienes hicieron la labor de lectura y escritura. Y estuvieron en observación en los procesos mencionados. Con la intención de ver como se relaciona el funcionamiento metacognitivo al escribir reseñas analíticas. En esta investigación hacen una revisión de diferentes fuentes bibliográficas respecto a la relación entre Metacognición y escritura, desde la perspectiva de diferentes autores, preocupados en como la Metacognición influye sobre los escritores dependiendo de su experiencia.^[33]

PERONARD Marianne (2005), presento el artículo “La Metacognición como herramienta didáctica”, resultados obtenidos en colaboración de FONDECYT, en el cual tuvo como objetivo conocer si existe una relación entre el conocimiento metacognitivo asociado al lenguaje escrito y las habilidades (lectura y escritura). El estudio fue realizado en estudiantes de segundo de secundaria, en el cual la investigación fue semi experimental con un grupo de control y dos grupos experimentales en los cuales hubo una intervención pedagógica. A estos grupos se les evaluó a través de cuatro pruebas, pre-test y post-test. Es decir antes y después de la intervención respectivamente.

Por otro lado se cita a la Metacognición como un constructo en el cuál no puede ser la excepción citar a Flavell (1979) el cual definió a la metacognición como “cognición de la cognición”. Y según el Diccionario de la Real Academia Española de la Lengua (RAE, 2001:581) el significado de la cognición como conocimiento (acción y efecto de conocer) un párrafo muy interesante es el siguiente: “Esos estudios pioneros, referidos a la memoria, buscaban descubrir cuánto sabían los niños acerca de la dificultad relativa de memorizar algo (proyectivo). Cuándo creían haberlo memorizado (retrospectivo) y que estrategias conocían para ayudarse a memorizar (introspectivo)” en el presente artículo muestra como este constructo ha sido tratado y enfocado por diferentes estudiosos. ^[34]

De la misma manera que los artículos mencionados anteriormente tenemos información muy diversa en el cual se estudia la Metacognición desde su significado, y su relación con otras variables.

En el proyecto PISA 2009 el término Metacognición no fue la excepción, específicamente la Metacognición en Lectura, esta se presenta como un constructo parte de las estrategias de aprendizaje. A su vez está formada por dos índices de la siguiente manera:



METASUM: Meta cognición – Resumen, esta variable ha sido estandarizada en base al promedio de la OECD, en base a las estrategias que utiliza el estudiante al resumir una lectura.

*“La adecuada utilización de estrategias ligadas al índice **resumir** por parte de los alumnos conlleva la consecución de niveles más altos de rendimiento en lectura”.*

(Fuente: PISA 2009 Informe Español)

Índice ‘resumir’

Tarea de lectura: Acabas de leer un texto de dos páginas, largo y bastante difícil, sobre las fluctuaciones del nivel del agua de un lago de África. Tienes que escribir un resumen.

¿Hasta qué punto consideras útiles las siguientes estrategias para escribir un resumen de este texto de dos páginas? (escala de seis posibles respuestas, desde ‘nada útil a ‘muy útil’)

- a) Escribo un resumen. Luego compruebo que el resumen recoge todos los párrafos, porque se debe incluir la información de cada párrafo.
- b) Intento copiar al pie de la letra todas las frases que puedo.
- c) Antes de escribir el resumen, leo el texto todas las veces posibles.
- d) Compruebo con atención si los hechos más importantes del texto están reflejados en el resumen.
- e) Leo todo el texto, subrayando las frases más importantes. Luego las reformulo con mis propias palabras a modo de resumen.

(Fuente: PISA 2009 Informe Español)

UNDREM: Meta cognición - Comprender y recordar, esta variable ha sido transformada ubicando al promedio de la OECD en cero y tiene una desviación 1. Se debe entender que al tener una mayor puntuación mejor es la estrategia para comprender y recordar que utiliza el estudiante. De la misma manera tener un valor positivo indica que está por encima del promedio de la OECD, por otro lado tener un valor negativo no indica una respuesta negativa más bien que se encuentra por debajo del promedio.

Índice 'comprender y recordar'

Tarea de lectura: Tienes que comprender y recordar la información de un texto.

¿Hasta qué punto consideras útiles las siguientes estrategias para entender y memorizar el texto? (escala de seis posibles respuestas, desde 'nada útil a 'muy útil').

- a) Me centro en las partes del texto que son fáciles de entender.
- b) Leo el texto rápidamente dos veces.
- c) Tras leer el texto, comento su contenido con otras personas.
- d) Subrayo las partes importantes del texto.
- e) Resumo el texto con mis propias palabras.
- f) Leo el texto en alto a otra persona.

(Fuente: PISA 2009 Informe Español)

ESCS: Índice de estatus económico, social y cultural, este índice fue construido para conocer el estatus de las familias de los estudiantes que participaron en el proyecto en base a las ocupaciones y nivel educativo de sus padres como el de las posesiones de su hogar y fueron denominadas respectivamente de la siguiente manera: HISEI, PARED y HOMEPOS.

HISEI: Máximo nivel ocupacional de los padres

PARED: Máximo nivel educativo de los padres.

HOMEPOS: Índice de posesiones en el hogar, a su vez este último índice se derivó de WEALTH, HEDRES y CULTPOSS.

WEALTH: Índice de Riqueza familiar.

HEDRES: Recursos educativos en el hogar.

CULTPOSS: Posesiones relacionadas con la cultura.

De la misma manera se cuentan con variables como JOYREAD que se refiere a Leer por placer, LMINS como Tiempo de aprendizaje, DISCLIMA como Ambiente disciplinario. Se han obtenido las variables PMAT5, PCL5, PCienc5 como promedios de otros cinco indicadores en las áreas de matemáticas, comprensión lectora y ciencias.

ZONA: Zona en la que se encuentra ubicado el colegio.

PROPQUAL: Proporción de profesores calificados.

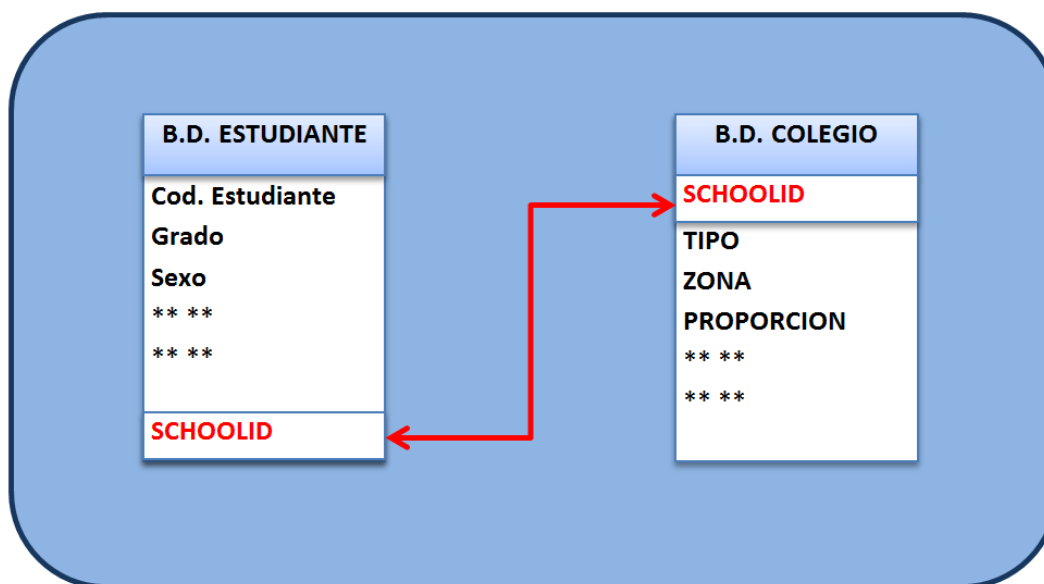
SCMATEDU: Recursos de calidad en el colegio.

COMPWEB: Proporción de computadoras conectadas a internet.

4.2 BASE DE DATOS PISA 2009-PERÚ

La unión de las bases de datos de los estudiantes y del colegio es posible debido a que tienen en común una variable SCHOOLID, que es el código del colegio, que vendría a ser como la llave que permite la integración.

UNIÓN DE LAS B.D. ESTUDIANTE Y COLEGIO

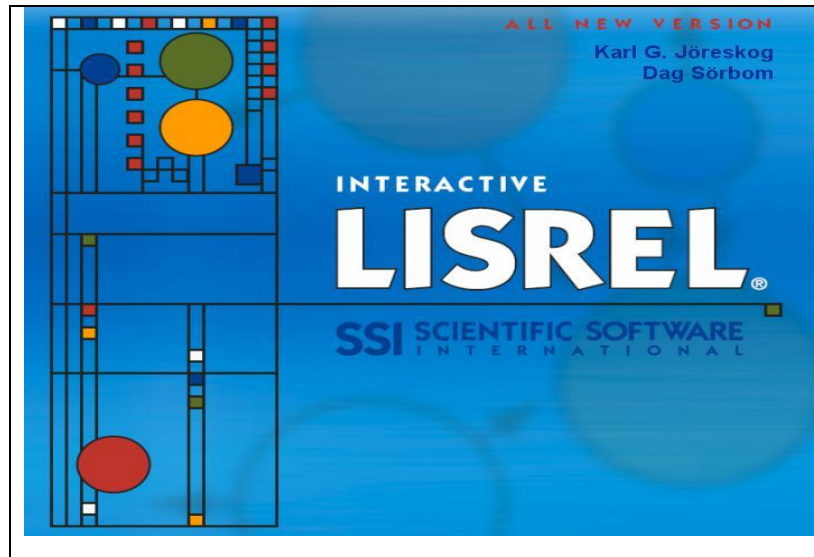


ECP1.sav [Conjunto_de_datos1] - IBM SPSS Statistics Editor de datos										
Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda										
5786 : XST01Q01 7										
	XWVARSTRR	XRANDUNIT	ZSC02Q01	ZSC04Q01	ZSTRATUM	ZZONA	ZTC	ZSC06Q01	ZSC06Q02	
1	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
2	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
3	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
4	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
5	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
6	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
7	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
8	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
9	47	2	Public	City	PER: Public/Urban	Urbana	Publico	14,00	20,	
10	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	
11	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	
12	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	
13	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	
14	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	
15	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	
16	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	
17	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	
18	26	1	Public	Town	PER: Public/Rural	Rural	Publico	534,00	478,	

ECP1.sav [Conjunto_de_datos1] - IBM SPSS Statistics Editor de datos										
Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda										
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	SCHOOLID	Cadena	5	0	School ID 5-digit	Ninguna	Ninguna	10	Izquierda	Nominal
2	StIDStd	Cadena	5	0	Student ID 5-digit	Ninguna	Ninguna	9	Izquierda	Nominal
3	XST01Q01	Numérico	2	0	Grade	{96, Ungrad...	96 - 99	9	Derecha	Ordinal
4	XST04Q01	Numérico	1	0	Sex	{1, Female}...	7, 8, 9	9	Derecha	Nominal
5	XMETASUM	Numérico	8	2	Meta-cognition: Summarising	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
6	XUNDREM	Numérico	8	2	Meta-cognition: Understanding and Remembering	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
7	XESCS	Numérico	8	2	Index of economic, social and cultural status (WLE)	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
8	XHISEI	Numérico	2	0	Highest parental occupational status	{97, N/A}...	97, 98, 99	12	Derecha	Escala
9	XPARED	Numérico	4	2	Highest parental education in years	{97,00, N/A}...	97,00, 98,0...	11	Derecha	Escala
10	XHOMEPOS	Numérico	8	2	Home Possessions	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
11	XWEALTH	Numérico	8	2	Wealth	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
12	XHEDRES	Numérico	8	2	Home educational resources	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
13	XCULTPOSS	Numérico	8	2	Cultural Possessions	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
14	XONLNREAD	Numérico	8	2	Online Reading	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
15	XJOYREAD	Numérico	8	2	Joy/Like Reading	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
16	XLMINS	Numérico	8	2	Learning time (minutes per week) - Test Language	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
17	XDISCLIMA	Numérico	8	2	Disciplinary climate	{9997,00, N/...	9997,00, 99...	12	Derecha	Escala
18	XPMat5	Numérico	8	2	Promedio de Matematicas	Ninguna	Ninguna	10	Derecha	Escala
19	XPCL5	Numérico	8	0	Promedio de Comprension Lectora	Ninguna	Ninguna	10	Derecha	Escala

4.3 ANÁLISIS MULTINIVEL EN LISREL

Los resultados obtenidos en la presente tesis se realizarán con el software LISREL, debido a que tiene implementado las técnicas para desarrollar los modelos de análisis multinivel en nuestro caso, los modelos no lineales, entre otras técnicas estadísticas que actualmente están tomando un mayor interés, en la página de Scientific Software International - SSI, se tiene información para la obtención del software. (Scientific Software International - SSI en la dirección <http://www.ssicentral.com/>). En su nueva versión se reconoce la autoría de Karl Jöreskog y de Dag Sorbom.



Captura de imagen de la página Scientific Software International, SSI

El software Linear Structural Relations (**LISREL**) ha ido actualizándose desde sus primeras versiones, alrededor del año 1975, este software es reconocido como uno de los primeros al utilizar los modelos de ecuaciones estructurales (Scientific Software International – SSI, <http://www.ssicentral.com/>), especialmente cuando los datos se han obtenido a partir de muestreos complejos. También es posible tener una interfaz gráfica. LISREL cuenta con la unión de dos módulos principales: LISREL, y PRESNIS, en el cuál esta unión ha fortalecido al software para las diferentes aplicaciones de la misma manera se han ido implementando más aplicaciones, presentamos una breve descripción de las aplicaciones que se pueden realizar:

LISREL: Para los modelos de ecuaciones estructurales.

PRESNIS: Para manipular datos y análisis estadísticos básicos.

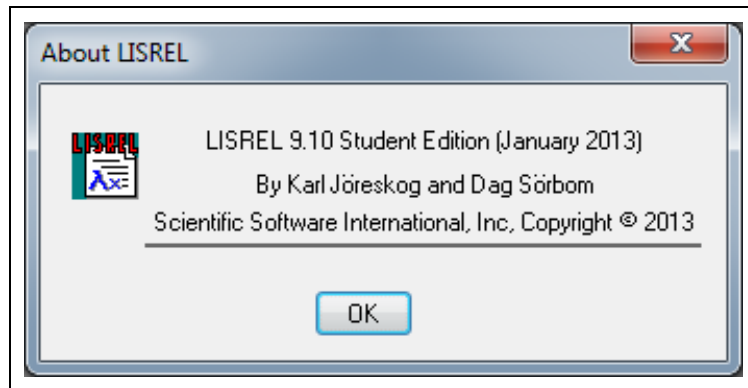
MULTILEV: Para modelos lineales y no lineales jerárquicos.

SURVEYGLIM: Para modelos lineales generalizados.

MAPGLIM: Para los modelos lineales generalizados para datos MULTINIVEL.

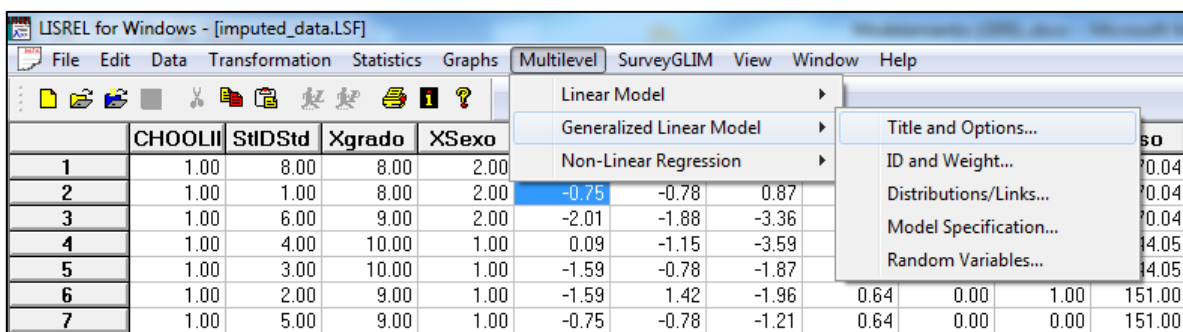
El software requiere de una licencia para ser utilizado y puede ser adquirido en la página citada anteriormente. sin embargo quienes desean iniciarse en estas técnicas de análisis pueden encontrarlo en su versión libre, es decir en la versión

de estudiante, claro está con algunas limitaciones. Como por ejemplo, existen restricciones respecto al número de variables, en este caso en particular, se tiene un máximo de 15 variables observadas a utilizar en los modelos de análisis multinivel como para los modelos de ecuaciones estructurales, por otro lado no hay un límite de días para su uso.

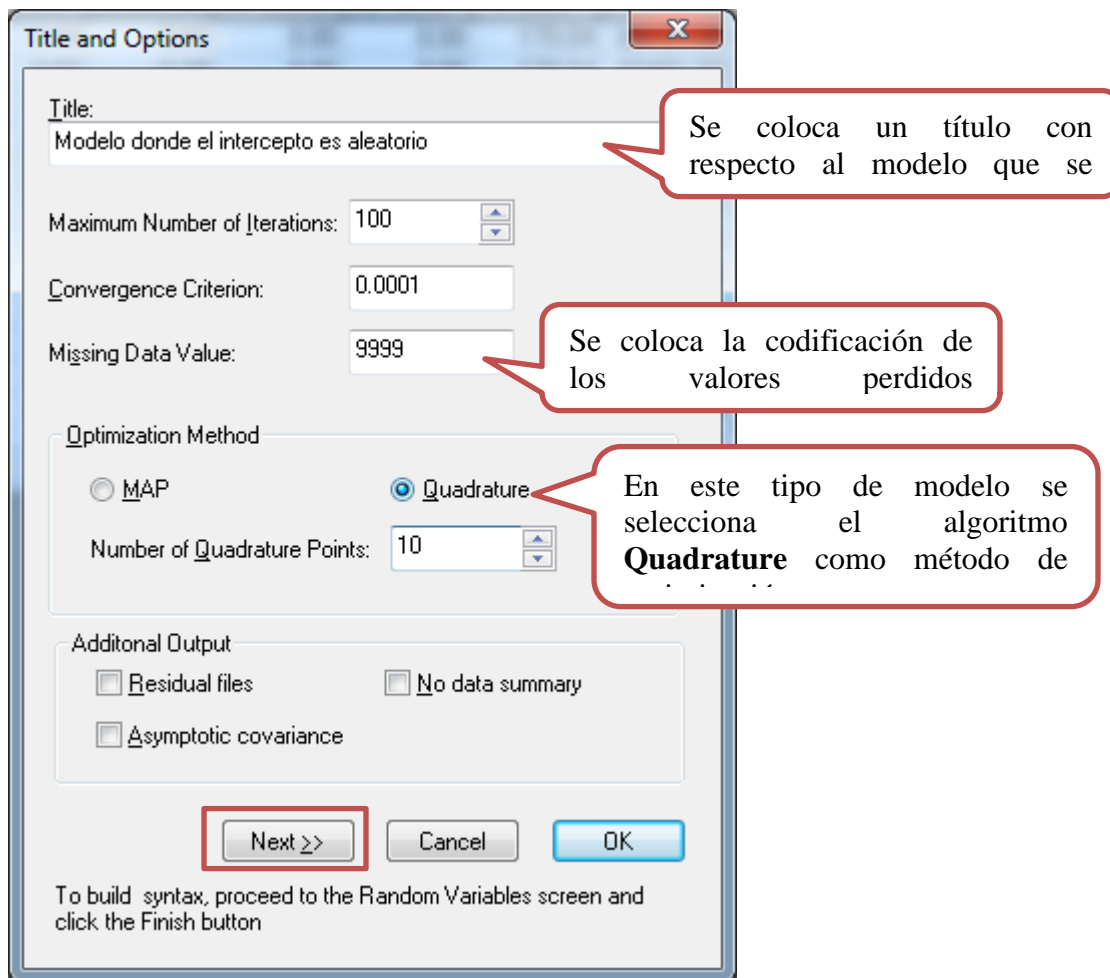


Modelo Logit

1. Se selecciona **Multilevel > Generalized Linear Model**



2. Aparecerá la siguiente ventana



3. En la ventana de la izquierda, se debe indicar las unidades del segundo nivel, así como los pesos muestrales. En la ventana de la derecha, se debe indicar la distribución y la función de enlace a usar.

ID and Weight Variables

Variables in data:

- SCHOOLID
- StDStd
- Xgrado
- XSexo
- XMETASUM
- XUNDREM
- XESCS
- XDISCLIM
- XPCL50
- XPCL5P
- Peso
- ZSTRATUM
- ZZONA
- ZTC
- ZPROPQUA
- ZSCMATED

Add >> << Remove

Level 2 ID variable: SCHOOLID

Add >> << Remove

Level 3 ID variable:

Add >> << Remove

Weight variable: Peso

<< Previous Next >> Cancel OK

To build syntax, proceed to the Random Variables screen and click the Finish button

Distributions and Links

Distribution type: Bernoulli

Link function: Logit

Include intercept? ☒ Yes ☐ No

Dispersion parameter ☐ Yes ☒ Fixed value:

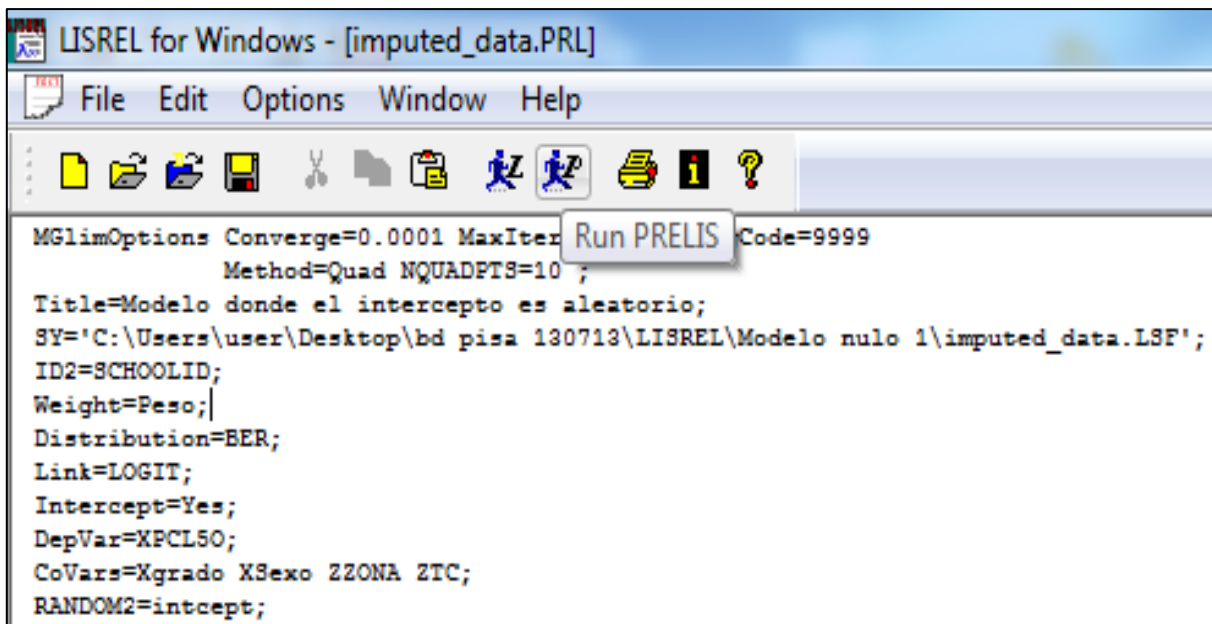
Estimate scale?

<< Previous Next >> Cancel OK

To build syntax, proceed to the Random Variables screen and click the Finish button

4. Se deberá indicar la variable dependiente y las variables independientes.
En el segundo nivel sólo el intercepto es aleatorio.

5. Presionamos el botón **FINISH**, obteniéndose la siguiente sintaxis:



The screenshot shows the LISREL for Windows application window titled "LISREL for Windows - [imputed_data.PRL]". The menu bar includes "File", "Edit", "Options", "Window", and "Help". Below the menu is a toolbar with various icons. The main command window contains the following syntax:

```
MGlimOptions Converge=0.0001 MaxIter Code=9999
Method=Quad NQUADPTS=10 ;
Title=Modelo donde el intercepto es aleatorio;
SY='C:\Users\user\Desktop\bd pisa 130713\LISREL\Modelo nulo 1\imputed_data.LSF';
ID2=SCHOOLID;
Weight=Peso;
Distribution=BER;
Link=LOGIT;
Intercept=Yes;
DepVar=XPCL50;
CoVars=Xgrado XSexo ZZONA ZTC;
RANDOM2=intcept;
```

6. Presionamos el botón **RUN PRELIS**, obteniéndose lo siguiente:

Modelo 1

o=====o Modelo donde el intercepto es aleatorio o=====o	
Model and Data Descriptions	
Sampling Distribution	= Bernoulli
Link Function	= Logistic
PROB(Success)= 1.0/[1.0+EXP(-ETA)]	
Level-1 Weight Variable	= Peso
Number of Level-2 Units	= 240
Number of Level-1 Units	= 5985

Como se aprecia en la salida anterior, la distribución empleada es la de Bernoulli con una función de enlace Logit. Y se indica cómo se calcularía la probabilidad de éxito. Se consideró los pesos muestrales de los estudiantes, la base de datos está conformada por 5985 estudiantes y 240 colegios, la cual corresponden al nivel uno y nivel dos respectivamente.

Inicialmente partimos sin considerar efectos aleatorios, entonces se tiene que el ajuste del modelo es de la siguiente forma:

o=====o Results for the model without any random effects o=====o			
Goodness of fit statistics			
Statistic	Value	DF	Ratio
-----	-----	--	-----
Likelihood Ratio Chi-square	33817.5076	5979	5.6560
Pearson Chi-square	301924.1381	5979	50.4974

Se muestran las estadísticas de máxima verosimilitud con un valor de 33817.5 con 5979 grados de libertad de la misma manera el coeficiente chi cuadrado de Pearson con 301924.1.

Considerando al intercepto como efecto aleatorio en el segundo nivel, tenemos que las medidas estadísticas para evaluar la adecuación del modelo son Desvianza con 2802.3, criterio de información de Akaike (AIC) con 2814.3 y el criterio bayesiano de schwarz (BIC) 2854.5 En particular si se desea verificar la capacidad predictiva del modelo (bondad de ajuste) nuestro estadístico es la desvianza $D = 2802.3$ la cual es inferior al punto crítico Chi cuadrado al 95% de nivel de confianza con 5978 grados de libertad $X^2_{(5978)} = 6158.98$, por lo tanto, no se rechaza la hipótesis nula, es decir el modelo ajustado si es significativo.

<pre> ===== Optimization Method: Adaptive Quadrature ===== </pre>	
Number of quadrature points =	10
Number of free parameters =	6
Number of iterations used =	3
-2lnL (deviance statistic) =	2802.29892
Akaike Information Criterion	2814.29892
Schwarz Criterion	2854.48099

La función de enlace logit será:

$$Prob(XPCL50 = 1) = \frac{1}{1 + e^{-\eta_{ij}}}$$

Modelo expresados en dos niveles serán:

Nivel 1:

$$\eta_{ij} = \beta_{0j} + \beta_{1j} \times XGrado_{ij} + \beta_{2j} \times XSexo_{ij} + \varepsilon_{ij}$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \times ZZona_j + \gamma_{02} \times ZTipo_j + \mu_{0j}$$

$$\beta_{kj} = \gamma_{k0}, \quad k = 1, 2$$

Otra forma de expresar es:

$$\eta_{ij} = \gamma_{00} + \gamma_{01} \times ZZona_j + \gamma_{02} \times ZTipo_j + \gamma_{10} \times XGrado_{ij} + \gamma_{20} \times XSexo_{ij} + \mu_{0j} + \varepsilon_{ij}$$

Los coeficientes estimados, considerando el efecto aleatorio en el intercepto son:

Estimated regression weights				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept	-9.5714	0.8939	-10.7071	0.0000
Xgrado	0.7769	0.0842	9.2299	0.0000
XSexo1	0.3731	0.1177	3.1701	0.0015
ZZONA1	-1.8477	0.4904	-3.7678	0.0002
ZTC1	-2.0648	0.2573	-8.0256	0.0000

Como se aprecia del cuadro los p_valores de los coeficientes estimados son significativos ($p_valor < 0.05$). El intercepto resulta ser -9.5714, es el valor promedio del predictor lineal cuya función de enlace es el logit. En la variable Grado su coeficiente estimado fue de 0.7769, en donde este valor variara debido al grado en que se encuentre cada alumno, esto quiere decir que a medida en que un estudiante aumenta en un grado se estima que el predictor lineal aumentará en promedio 0.7769 unidades. En la variable Sexo su coeficiente estimado fue de 0.3731 este será aplicado cuando el sexo sea femenino. En la variable Zona su coeficiente estimado fue de -1.8477, y será aplicado cuando la zona del colegio sea rural. En la variable Tipo de colegio su coeficiente estimado fue de -2.0648, y será aplicado cuando el colegio sea público. (Observación: inicialmente las categorías para variable sexo fue 0: femenino y 1: masculino, sin embargo cuando uno ingresa la variable como categórica internamente el software genera una transformación respecto a la última categoría (masculino) como una especie de variable dummy entonces haciendo la comparación respecto a la categoría masculino tomará el valor cero y para el caso femenino 1. Por eso podemos notar que en el reporte aparece la variable XSEXO1 llamándose originalmente XSEXO,

en adelante se realizaran las interpretaciones tomando en cuenta esta observación.)

Entonces el modelo estimado será:

$$\hat{\eta}_{ij} = -9.5714 + 0.7769 \times XGrado_{ij} + 0.3731 \times XSexo_{ij} - 1.8477 \times ZZona_j - 2.0648 \times ZTipo_j$$

Las probabilidades asociadas a obtener o superar el promedio de la OECD en comprensión lectora, para un estudiante de distinto grado y sexo, el tipo de colegio en el que estudie y la zona en la que se ubique el colegio, son presentadas en la siguiente tabla:

**Cuadro N°4.1 : PROBABILIDADES CORRESPONDIENTES AL MODELO
LOGIT 1**

Grupo	Código	η	Probabilidad (η)
1S, Femenino, Rural, Publico	(7 , 1 , 1 , 1)	-7.6724	0.05%
1S, Femenino, Rural, Privado	(7 , 1 , 1 , 0)	-5.6076	0.37%
1S, Femenino, Urbana, Publico	(7 , 1 , 0 , 1)	-5.8247	0.29%
1S, Femenino, Urbana, Privado	(7 , 1 , 0 , 0)	-3.7599	2.28%
1S, Masculino, Rural, Publico	(7 , 0 , 1 , 1)	-8.0455	0.03%
1S, Masculino, Rural, Privado	(7 , 0 , 1 , 0)	-5.9807	0.25%
1S, Masculino, Urbana, Publico	(7 , 0 , 0 , 1)	-6.1978	0.20%
1S, Masculino, Urbana, Privado	(7 , 0 , 0 , 0)	-4.133	1.58%
2S, Femenino, Rural, Publico	(8 , 1 , 1 , 1)	-6.8955	0.10%
2S, Femenino, Rural, Privado	(8 , 1 , 1 , 0)	-4.8307	0.79%
2S, Femenino, Urbana, Publico	(8 , 1 , 0 , 1)	-5.0478	0.64%
2S, Femenino, Urbana, Privado	(8 , 1 , 0 , 0)	-2.983	4.82%
2S, Masculino, Rural, Publico	(8 , 0 , 1 , 1)	-7.2686	0.07%
2S, Masculino, Rural, Privado	(8 , 0 , 1 , 0)	-5.2038	0.55%
2S, Masculino, Urbana, Publico	(8 , 0 , 0 , 1)	-5.4209	0.44%
2S, Masculino, Urbana, Privado	(8 , 0 , 0 , 0)	-3.3561	3.37%
3S, Femenino, Rural, Publico	(9 , 1 , 1 , 1)	-6.1186	0.22%
3S, Femenino, Rural, Privado	(9 , 1 , 1 , 0)	-4.0538	1.71%
3S, Femenino, Urbana, Publico	(9 , 1 , 0 , 1)	-4.2709	1.38%
3S, Femenino, Urbana, Privado	(9 , 1 , 0 , 0)	-2.2061	9.92%
3S, Masculino, Rural, Publico	(9 , 0 , 1 , 1)	-6.4917	0.15%
3S, Masculino, Rural, Privado	(9 , 0 , 1 , 0)	-4.4269	1.18%
3S, Masculino, Urbana, Publico	(9 , 0 , 0 , 1)	-4.644	0.95%
3S, Masculino, Urbana, Privado	(9 , 0 , 0 , 0)	-2.5792	7.05%
4S, Femenino, Rural, Publico	(10 , 1 , 1 , 1)	-5.3417	0.48%
4S, Femenino, Rural, Privado	(10 , 1 , 1 , 0)	-3.2769	3.64%
4S, Femenino, Urbana, Publico	(10 , 1 , 0 , 1)	-3.494	2.95%
4S, Femenino, Urbana, Privado	(10 , 1 , 0 , 0)	-1.4292	19.32%
4S, Masculino, Rural, Publico	(10 , 0 , 1 , 1)	-5.7148	0.33%
4S, Masculino, Rural, Privado	(10 , 0 , 1 , 0)	-3.65	2.53%
4S, Masculino, Urbana, Publico	(10 , 0 , 0 , 1)	-3.8671	2.05%
4S, Masculino, Urbana, Privado	(10 , 0 , 0 , 0)	-1.8023	14.16%
5S, Femenino, Rural, Publico	(11 , 1 , 1 , 1)	-4.5648	1.03%
5S, Femenino, Rural, Privado	(11 , 1 , 1 , 0)	-2.5	7.59%

5S, Femenino, Urbana, Publico	(11 , 1 , 0 , 1)	-2.7171	6.20%
5S, Femenino, Urbana, Privado	(11 , 1 , 0 , 0)	-0.6523	34.25%
5S, Masculino, Rural, Publico	(11 , 0 , 1 , 1)	-4.9379	0.71%
5S, Masculino, Rural, Privado	(11 , 0 , 1 , 0)	-2.8731	5.35%
5S, Masculino, Urbana, Publico	(11 , 0 , 0 , 1)	-3.0902	4.35%
5S, Masculino, Urbana, Privado	(11 , 0 , 0 , 0)	-1.0254	26.40%

Estas probabilidades fueron calculadas por el autor de la presente tesis

Como se puede apreciar en la tabla, se obtuvo que la probabilidad de que un estudiante supere o iguale al promedio de la OECD dado que se encuentra en el quinto de secundaria, sea de género masculino, pertenezca a la zona urbana y provenga de un colegio privado fue de 26.40%.

La probabilidad de que un estudiante supere o iguale al promedio de la OECD dado que se encuentra en el quinto de secundaria, sea de género femenino, pertenezca a la zona urbana y provenga de un colegio privado fue de 34.25%.

De manera similar en los casos anteriores pero considerando que pertenecen al cuarto de secundaria las probabilidades de éxito fueron de 14.16% y 19.32% para el género masculino y femenino respectivamente. Se puede apreciar que en el área de comprensión lectora las mujeres tienen una mayor probabilidad de éxito que los varones y que a medida que los estudiantes van aumentando de grado estas probabilidades tienden a aumentar.

Odds Ratio and 95% Odds Ratio Confidence Intervals				
Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
-----	-----	-----	-----	-----
intcept	-9.5714	0.0001	0.0000	0.0004
Xgrado	0.7769	2.1746	1.8439	2.5647
XSexo1	0.3731	1.4523	1.1531	1.8291
ZZONA1	-1.8477	0.1576	0.0603	0.4121
ZTC1	-2.0648	0.1268	0.0766	0.2100

Analizando los intervalos para los odds ratio se puede ver que el grado, el sexo, la zona y el tipo de colegio se encuentran relacionados cada uno de ellos con la variable respuesta, condición de la comprensión lectora según el promedio OECD. Un estudiante que se encuentra en determinado grado tiene 2.17 veces de

posibilidad de obtener o superar el promedio de la OECD en comprensión lectora respecto a que se encuentre en un grado anterior. Por otro lado las mujeres tienen 1.45 veces de posibilidad aproximadamente de obtener o superar el promedio de la OECD en comprensión lectora con respecto a los varones.

Además se puede observar que para el caso que un estudiante pertenezca a la zona rural o provenga de un colegio público las posibilidades de éxito son inferiores comparados con el hecho de que pertenezca a la zona urbana o provenga de un colegio privado.

El siguiente cuadro muestra la estimación de la varianza en el segundo nivel, y como se puede apreciar el $p_valor = 0.000 < 0.05$, lo que indicaría que la probabilidad de obtener puntajes mayores o iguales a los de la OECD difieren de colegio en colegio.

Estimated level 2 variances and covariances				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept/intcept	1.7445	0.3043	5.7332	0.0000

En la siguiente tabla se muestra el coeficiente de correlación intraclase del modelo, el cual resulta ser 0.347, esto indica que el 34.7% de la variabilidad es explicada por las unidades del segundo nivel (colegios).

Calculation of the intracluster correlation	

residual variance = $\pi * \pi / 3$ (assumed)	
cluster variance = 1.7445	
intracluster correlation = $1.7445 / (1.7445 + (\pi * \pi / 3))$	= 0.347

MODELO FINAL LOGIT

```
o=====o
| Modelo Final |
o=====o

Model and Data Descriptions

Sampling Distribution          = Bernoulli
Link Function                 = Logistic
PROB(Success)= 1.0/[1.0+EXP(-ETA)]

Level-1 Weight Variable      = Peso
Number of Level-2 Units      = 240
Number of Level-1 Units      = 5985
```

Como podemos apreciar de esta salida veremos que la distribución empleada fue la de Bernoulli con función de enlace Logit, además nos indica cómo se calcularía la probabilidad de éxito. Se consideró los pesos muestrales de los estudiantes, la data cuenta con 5985 estudiantes y 240 colegios.

Sin considerar efectos aleatorios, el ajuste del modelo es de la siguiente forma:

```
o=====o
| Results for the model without any random effects |
o=====o

Goodness of fit statistics

Statistic          Value          DF          Ratio
-----
Likelihood Ratio Chi-square      42744.2472      5974      7.1550
Pearson Chi-square      1305124.3031      5974      218.4674
```

Considerando al intercepto como efecto aleatorio en el segundo nivel, tenemos que las medidas estadísticas para evaluar la adecuación del modelo son Desvianza con 2469.2, criterio de información de Akaike (AIC) con 2491.2 y el criterio bayesiano de schwarz (BIC) 2564.8 En particular si se desea verificar la

capacidad predictiva del modelo (bondad de ajuste) nuestro estadístico es la desviación $D = 2469.2$ la cual es inferior al punto crítico Chi cuadrado al 95% de nivel de confianza con 5973 grados de libertad $X^2_{(5973)} = 6153.91$, con lo tanto no se rechaza la hipótesis nula, es decir el modelo ajustado si es significativo.

=====o	
Optimization Method: Adaptive Quadrature	
=====o	
Number of quadrature points =	10
Number of free parameters =	11
Number of iterations used =	3
-2lnL (deviance statistic) =	2469.17025
Akaike Information Criterion	2491.17025
Schwarz Criterion	2564.83738

Los coeficientes estimados, considerando el efecto aleatorio en el intercepto es:

Estimated regression weights				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept	-8.3819	0.9542	-8.7838	0.0000
Xgrado	0.6015	0.0882	6.8204	0.0000
XSexo1	0.2977	0.1205	2.4711	0.0135
XMETASUM	0.7010	0.0776	9.0336	0.0000
XUNDREM	0.4318	0.0642	6.7315	0.0000
XESCS	0.4728	0.0642	7.3586	0.0000
ZZONA1	-0.6000	0.4241	-1.4149	0.1571
ZTC1	-0.6578	0.2150	-3.0589	0.0022
ZPROPQUA	1.2366	0.3448	3.5859	0.0003
ZSCMATED	0.3871	0.0888	4.3580	0.0000

Como se aprecia del cuadro la mayor parte de los p_valores de los coeficientes estimados son significativos ($p_valor < 0.05$). El intercepto resulta ser -8.3819, es el valor promedio del predictor lineal cuya función de enlace es el logit. En la variable Grado su coeficiente estimado fue de 0.6015, en donde este valor variara debido al grado en que se encuentre cada alumno. En la variable Sexo su coeficiente estimado fue de 0.2977, y será aplicado cuando el sexo sea femenino. En la variable METASUM su coeficiente estimado fue de 0.7010, en donde este valor variara debido al puntaje que obtenga cada alumno, es decir por cada unidad

que se incremente en la Metacognición en Lectura (Resumir) se estima que el predictor lineal aumentará en promedio 0.7010 unidades. En la variable UNDREM su coeficiente estimado fue de 0.4318, en donde este valor variara debido al puntaje que obtenga cada alumno, es decir por cada unidad en que se incremente en la Metacognición en Lectura (comprender y recordar) se estima que el predictor lineal aumentará en promedio en 0.4318 unidades. En la variable ESCS su coeficiente estimado fue de 0.4728, en donde este valor variara debido al puntaje que obtenga cada alumno, es decir por cada unidad que se incremente en el Índice de estatus económico, social y cultural se estima que el predictor lineal aumentara en promedio 0.4728 unidades. En la variable Zona su coeficiente estimado fue de - 0.6, y será aplicado cuando la zona del colegio sea rural. En la variable Tipo de colegio su coeficiente estimado fue de -0.6578, y será aplicado cuando el colegio sea público. En la variable PROPQUA su coeficiente estimado fue de 1.2366, es decir por cada unidad que se incremente sobre la proporción de docentes calificados en cada colegio se estima que el predictor lineal se incrementara en promedio en 1.2366 unidades, de manera similar para la variable SCMATD calidad de recursos educativos, donde su coeficiente estimado fue de 0.3871.

Entonces el modelo estimado será:

$$\begin{aligned}\hat{\eta}_{ij} = & -8.3819 + 0.6015 \times XGrado_{ij} + 0.2977 \times XSexo_{ij} + 0.7010 \times XMetasum_{ij} \\ & + 0.4318 \times XUndrem_{ij} + 0.4728 \times XESCS_{ij} - 0.6000 \times ZZona_j \\ & - 0.6578 \times ZTipo_j + 1.2366 \times ZPropqua_j + 0.3871 \times ZSCMated_j\end{aligned}$$

Las probabilidades asociadas a obtener o superar el promedio de la OECD en comprensión lectora, para un estudiante de un determinado grado, sexo, que haya obtenido un puntaje promedio en METASUM, UNDREM y en ESCS, y en el segundo nivel, el tipo de colegio, la zona en la que se ubique el colegio, en particular con un 50% de docentes calificados (*para el cálculo de la probabilidad se tomó en cuenta el valor 0.5 por el coeficiente 1.2366*) y que haya obtenido un

puntaje promedio en la calidad de recursos educativos del colegio (*por estar la variable estandarizada, el valor promedio se ubicaría en cero, por tanto para el cálculo de la probabilidad el coeficiente no afectaría, es por ello que se muestran solo las columnas para algunas variables*), son presentadas en la siguiente tabla:

**Cuadro N°4.2: PROBABILIDADES CORRESPONDIENTES AL MODELO
LOGIT FINAL**

Grupo	Código	η	Probabilidad (η)
1S, Femenino, Rural, Publico	(7 , 1 , 1 , 1)	-4.5132	1.08%
1S, Femenino, Rural, Privado	(7 , 1 , 1 , 0)	-3.8554	2.07%
1S, Femenino, Urbana, Publico	(7 , 1 , 0 , 1)	-3.9132	1.96%
1S, Femenino, Urbana, Privado	(7 , 1 , 0 , 0)	-3.2554	3.71%
1S, Masculino, Rural, Publico	(7 , 0 , 1 , 1)	-4.8109	0.81%
1S, Masculino, Rural, Privado	(7 , 0 , 1 , 0)	-4.1531	1.55%
1S, Masculino, Urbana, Publico	(7 , 0 , 0 , 1)	-4.2109	1.46%
1S, Masculino, Urbana, Privado	(7 , 0 , 0 , 0)	-3.5531	2.78%
2S, Femenino, Rural, Publico	(8 , 1 , 1 , 1)	-3.9117	1.96%
2S, Femenino, Rural, Privado	(8 , 1 , 1 , 0)	-3.2539	3.72%
2S, Femenino, Urbana, Publico	(8 , 1 , 0 , 1)	-3.3117	3.52%
2S, Femenino, Urbana, Privado	(8 , 1 , 0 , 0)	-2.6539	6.57%
2S, Masculino, Rural, Publico	(8 , 0 , 1 , 1)	-4.2094	1.46%
2S, Masculino, Rural, Privado	(8 , 0 , 1 , 0)	-3.5516	2.79%
2S, Masculino, Urbana, Publico	(8 , 0 , 0 , 1)	-3.6094	2.64%
2S, Masculino, Urbana, Privado	(8 , 0 , 0 , 0)	-2.9516	4.97%
3S, Femenino, Rural, Publico	(9 , 1 , 1 , 1)	-3.3102	3.52%
3S, Femenino, Rural, Privado	(9 , 1 , 1 , 0)	-2.6524	6.58%
3S, Femenino, Urbana, Publico	(9 , 1 , 0 , 1)	-2.7102	6.24%
3S, Femenino, Urbana, Privado	(9 , 1 , 0 , 0)	-2.0524	11.38%
3S, Masculino, Rural, Publico	(9 , 0 , 1 , 1)	-3.6079	2.64%
3S, Masculino, Rural, Privado	(9 , 0 , 1 , 0)	-2.9501	4.97%
3S, Masculino, Urbana, Publico	(9 , 0 , 0 , 1)	-3.0079	4.71%
3S, Masculino, Urbana, Privado	(9 , 0 , 0 , 0)	-2.3501	8.71%
4S, Femenino, Rural, Publico	(10 , 1 , 1 , 1)	-2.7087	6.25%
4S, Femenino, Rural, Privado	(10 , 1 , 1 , 0)	-2.0509	11.40%
4S, Femenino, Urbana, Publico	(10 , 1 , 0 , 1)	-2.1087	10.83%
4S, Femenino, Urbana, Privado	(10 , 1 , 0 , 0)	-1.4509	18.99%
4S, Masculino, Rural, Publico	(10 , 0 , 1 , 1)	-3.0064	4.71%
4S, Masculino, Rural, Privado	(10 , 0 , 1 , 0)	-2.3486	8.72%
4S, Masculino, Urbana, Publico	(10 , 0 , 0 , 1)	-2.4064	8.27%
4S, Masculino, Urbana, Privado	(10 , 0 , 0 , 0)	-1.7486	14.82%
5S, Femenino, Rural, Publico	(11 , 1 , 1 , 1)	-2.1072	10.84%
5S, Femenino, Rural, Privado	(11 , 1 , 1 , 0)	-1.4494	19.01%
5S, Femenino, Urbana, Publico	(11 , 1 , 0 , 1)	-1.5072	18.14%
5S, Femenino, Urbana, Privado	(11 , 1 , 0 , 0)	-0.8494	29.96%
5S, Masculino, Rural, Publico	(11 , 0 , 1 , 1)	-2.4049	8.28%
5S, Masculino, Rural, Privado	(11 , 0 , 1 , 0)	-1.7471	14.84%
5S, Masculino, Urbana, Publico	(11 , 0 , 0 , 1)	-1.8049	14.13%
5S, Masculino, Urbana, Privado	(11 , 0 , 0 , 0)	-1.1471	24.10%

Estas probabilidades fueron calculadas por el autor de la presente tesis

De la tabla anterior, dado que las variables METASUM, UNDREM, ESCS y SCMATED se encuentran estandarizadas, entonces sus promedios son el valor cero.

Odds Ratio and 95% Odds Ratio Confidence Intervals				
Parameter	Estimate	Odds Ratio	Bounds	
-----	-----	-----	-----	-----
intcept	-8.3819	0.0002	0.0000	0.0015
Xgrado	0.6015	1.8248	1.5351	2.1691
XSexo1	0.2977	1.3468	1.0635	1.7055
XMETASUM	0.7010	2.0158	1.7313	2.3469
XUNDREM	0.4318	1.5401	1.3581	1.7464
XESCS	0.4728	1.6045	1.4146	1.8198
ZZONA1	-0.6000	0.5488	0.2390	1.2601
ZTC1	-0.6578	0.5180	0.3398	0.7895
ZPROPQUA	1.2366	3.4438	1.7519	6.7698
ZSCMATED	0.3871	1.4727	1.2374	1.7528

Analizando los intervalos para los odds ratio se puede ver que las variables grado, sexo, Metacognición resumir, Metacognición comprender y recordar, índice de estatus económico social y cultural como el tipo de colegio, proporción de profesores calificados y materiales educativos de calidad si están relacionados con la variable dependiente puntaje (condición) en comprensión lectora según el promedio OECD (0: debajo del promedio OECD, 1: Obtenga o supere el promedio OECD). Sin embargo en este modelo en el cuál se han incrementado más variables la variable zona de colegio no está relacionada con la variable dependiente. Un estudiante que se encuentra en determinado grado tiene 1.82 veces de posibilidad de obtener o superar el promedio de la OECD en comprensión lectora respecto a que se encuentre en un grado anterior. Por otro lado las mujeres tienen 1.35 veces de posibilidad aproximadamente de obtener o superar el promedio de la OECD en comprensión lectora con respecto a los varones.

El siguiente cuadro muestra la estimación de la varianza en el segundo nivel, y como se puede apreciar el $p_valor = 0.000 < 0.05$, lo que indicaría que la

probabilidad de obtener puntajes mayores o iguales a los de la OECD difieren de colegio en colegio.

Estimated level 2 variances and covariances				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept/intcept	0.6534	0.1462	4.4691	0.0000

En la siguiente tabla se muestra el coeficiente intraclase del modelo, el cual resulta ser 0.166, esto indica que el 16.6% de la variabilidad es explicada por las unidades del segundo nivel (ver página 24[12]).

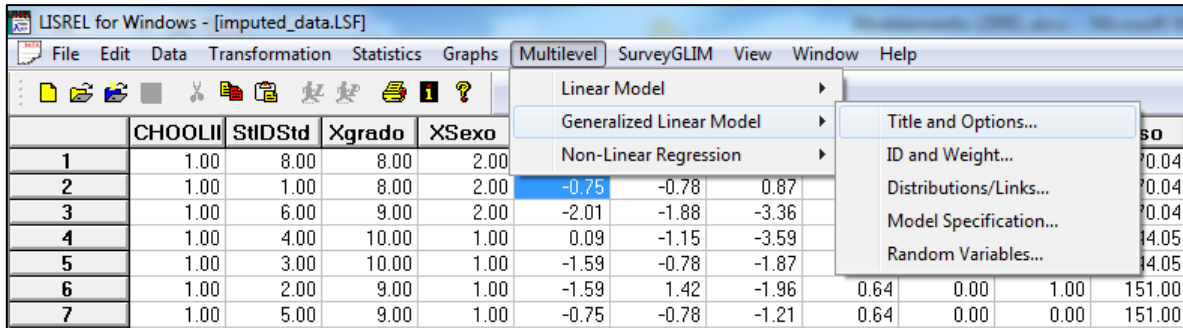
Calculation of the intracluster correlation

residual variance = $\pi^2 / 3$ (assumed)
cluster variance = 0.6534
intracluster correlation = $0.6534 / (0.6534 + (\pi^2 / 3)) = 0.166$

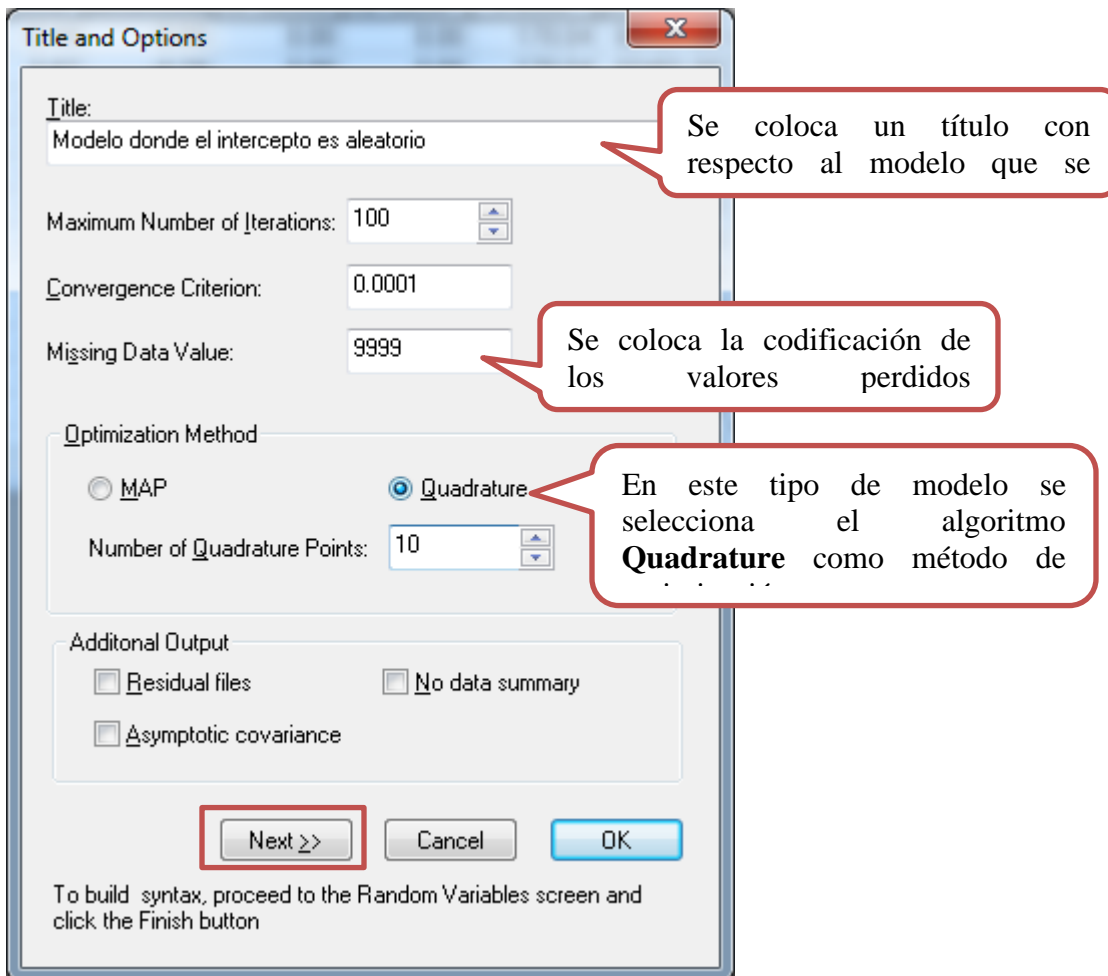
ANÁLISIS MULTINIVEL EN LISREL

Modelo Probit

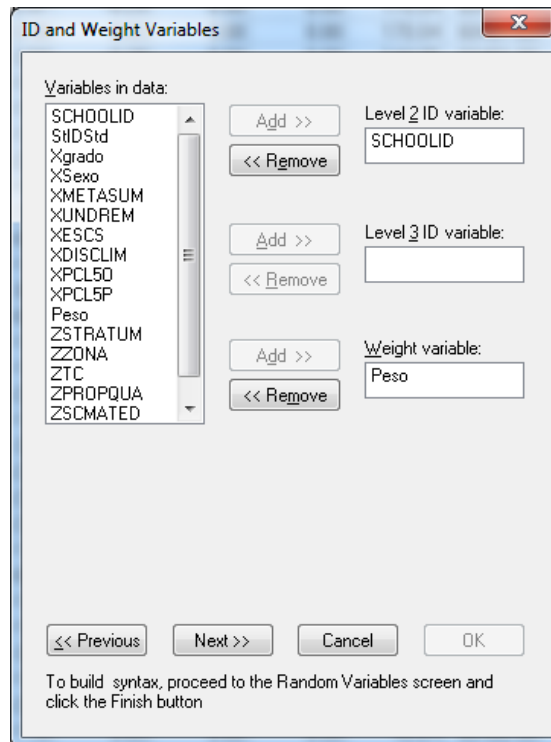
1. Se selecciona **Multilevel > Generalized Linear Model**



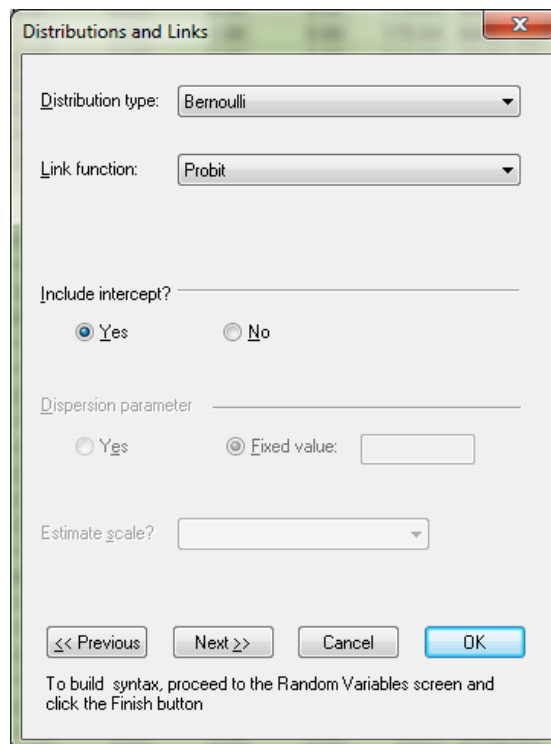
2. Aparecerá la siguiente ventana



3. En la ventana de la izquierda, se debe indicar las unidades del segundo nivel, así como los pesos muestrales. En la ventana de la derecha, se debe indicar la distribución y la función de enlace a usar.



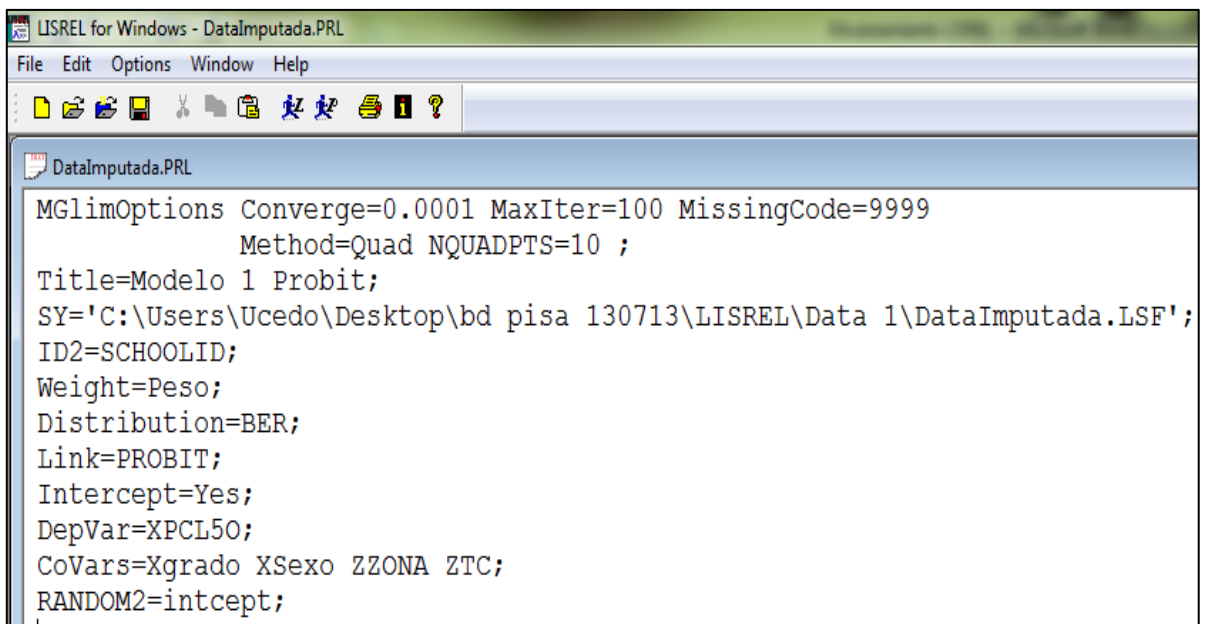
The 'ID and Weight Variables' dialog box is shown. It features a list of variables on the left: SCHOOLID, SIDSID, Xgrado, XSexo, XMETASUM, XUNDREM, XESCS, XDISCLIM, XPCL50, XPCL5P, Peso, ZSTRATUM, ZZONA, ZTC, ZPROPQUA, and ZSCMATED. To the right of the list are 'Add >>' and '<< Remove' buttons. Further right are fields for 'Level 2 ID variable:' (containing 'SCHOOLID'), 'Level 3 ID variable:' (empty), and 'Weight variable:' (containing 'Peso'). Each field has its own 'Add >>' and '<< Remove' buttons. At the bottom are 'Previous', 'Next', 'Cancel', and 'OK' buttons. A note at the bottom states: 'To build syntax, proceed to the Random Variables screen and click the Finish button.'



The 'Distributions and Links' dialog box is shown. It has a 'Distribution type:' dropdown menu set to 'Bernoulli'. Below it is a 'Link function:' dropdown menu set to 'Probit'. There is an 'Include intercept?' section with radio buttons for 'Yes' (selected) and 'No'. Below that is a 'Dispersion parameter' section with radio buttons for 'Yes' and 'Fixed value:' (selected), followed by an empty text box. At the bottom is an 'Estimate scale?' dropdown menu. At the very bottom are 'Previous', 'Next', 'Cancel', and 'OK' buttons. A note at the bottom states: 'To build syntax, proceed to the Random Variables screen and click the Finish button.'

4. Se deberá indicar la variable dependiente y las variables independientes.
En el segundo nivel sólo el intercepto es aleatorio.

5. Presionamos el botón **FINISH**, obteniéndose la siguiente sintaxis:



```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=9999
                Method=Quad NQUADPTS=10 ;
Title=Modelo 1 Probit;
SY='C:\Users\Ucedo\Desktop\bd pisa 130713\LISREL\Data 1\DataImputada.LSF';
ID2=SCHOOLID;
Weight=Peso;
Distribution=BER;
Link=PROBIT;
Intercept=Yes;
DepVar=XPCL50;
CoVars=Xgrado XSexo ZZONA ZTC;
RANDOM2=intcept;
```

6. Presionamos el botón **RUN PRELIS**, obteniéndose lo siguiente:

Modelo 1 - PROBIT

o=====o	
Modelo donde el intercepto es aleatorio	
o=====o	
Model and Data Descriptions	
Sampling Distribution	= Bernoulli
Link Function	= Probit
PROB(Success)= PHI(ETA)	
PHI denotes the CDF of the standard Normal distribution	
Level-1 Weight Variable	= Peso
Number of Level-2 Units	= 240
Number of Level-1 Units	= 5985

Como podemos apreciar de esta salida veremos que la distribución empleada fue la de Bernoulli con función de enlace Probit, además nos indica cómo se calcularía la probabilidad de éxito. Se consideró los pesos muestrales de los estudiantes, la data cuenta con 5985 estudiantes y 240 colegios.

Sin considerar efectos aleatorios, el ajuste del modelo es de la siguiente forma:

o=====o			
Results for the model without any random effects			
o=====o			
Goodness of fit statistics			
Statistic	Value	DF	Ratio
-----	-----	--	-----
Likelihood Ratio Chi-square	35552.7859	5979	5.9463
Pearson Chi-square	847528.7148	5979	141.7509

Considerando al intercepto como efecto aleatorio en el segundo nivel, tenemos que las medidas estadísticas para evaluar la adecuación del modelo son Desvianza con 2798.14, criterio de información de Akaike (AIC) con 2810.14 y el criterio bayesiano de schwarz (BIC) 2850.32 En particular si se desea verificar la capacidad predictiva del modelo (bondad de ajuste) nuestro estadístico es la desvianza $D = 2798.14$ la cual es inferior al punto crítico Chi cuadrado al 95% de

nivel de confianza con 5978 grados de libertad $X^2_{(5978)} = 6158.98$, con lo tanto no se rechaza la hipótesis nula, es decir el modelo ajustado si es significativo.

<pre> =====o Optimization Method: Adaptive Quadrature =====o </pre>	
Number of quadrature points =	10
Number of free parameters =	6
Number of iterations used =	3
-2lnL (deviance statistic) =	2798.14251
Akaike Information Criterion	2810.14251
Schwarz Criterion	2850.32457

La función de enlace probit será:

$$Prob(XPCL50 = 1) = \Phi^{-1}(\eta_{ij})$$

Modelo expresados en dos niveles serán:

Nivel 1:

$$\eta_{ij} = \beta_{0j} + \beta_{1j} \times XGrado_{ij} + \beta_{2j} \times XSexo_{ij} + \varepsilon_{ij}$$

Nivel 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} \times ZZona_j + \gamma_{02} \times ZTipo_j + \mu_{0j}$$

$$\beta_{kj} = \gamma_{k0}, \quad k = 1, 2$$

Otra forma de expresar es:

$$\eta_{ij} = \gamma_{00} + \gamma_{01} \times ZZona_j + \gamma_{02} \times ZTipo_j + \gamma_{10} \times XGrado_{ij} + \gamma_{20} \times XSexo_{ij} + \mu_{0j} + \varepsilon_{ij}$$

Los coeficientes estimados, considerando el efecto aleatorio en el intercepto es:

Estimated regression weights				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept	-5.1356	0.4632	-11.0882	0.0000
Xgrado	0.4125	0.0437	9.4325	0.0000
XSexo1	0.1982	0.0628	3.1584	0.0016
ZZONA1	-0.9065	0.2392	-3.7902	0.0002
ZTC1	-1.0910	0.1335	-8.1703	0.0000

Como se aprecia del cuadro los p_valores de los coeficientes estimados son significativos ($p_valor < 0.05$). El intercepto resulta ser -5.1356, es el promedio del probit. En la variable Grado su coeficiente estimado fue de 0.4125, en donde este valor variara debido al grado en que se encuentre cada alumno, es decir a medida en que un estudiante avance en un grado se estima que el predictor lineal aumentará en promedio 0.4125 unidades. En la variable Sexo, el coeficiente estimado fue de 0.1982, y será aplicado cuando el sexo sea femenino. En la variable Zona su coeficiente estimado fue de -0.9065, y será aplicado cuando la zona del colegio sea rural. En la variable Tipo de colegio su coeficiente estimado fue de -1.091, y será aplicado cuando el colegio sea público.

Entonces el modelo estimado será:

$$\hat{\eta}_{ij} = -5.1356 + 0.4125 \times XGrado_{ij} + 0.1982 \times XSexo_{ij} - 0.9065 \times ZZona_j - 1.0910 \times ZTipo_j$$

Las probabilidades asociadas a obtener o superar el promedio de la OECD en comprensión lectora, para un estudiante de distinto grado y sexo, el tipo de colegio

en el que estudie y la zona en la que se ubique el colegio, son presentadas en la siguiente tabla:

**Cuadro N°4.3: PROBABILIDADES CORRESPONDIENTES AL
MODELO PROBIT 1**

Grupo	Código	η	Probabilidad (η)
1S, Femenino, Rural, Publico	(7 , 1 , 1 , 1)	-4.0474	0.00%
1S, Femenino, R.302ural, Privado	(7 , 1 , 1 , 0)	-2.9564	0.16%
1S, Femenino, Urbana, Publico	(7 , 1 , 0 , 1)	-3.1409	0.08%
1S, Femenino, Urbana, Privado	(7 , 1 , 0 , 0)	-2.0499	2.02%
1S, Masculino, Rural, Publico	(7 , 0 , 1 , 1)	-4.2456	0.00%
1S, Masculino, Rural, Privado	(7 , 0 , 1 , 0)	-3.1546	0.08%
1S, Masculino, Urbana, Publico	(7 , 0 , 0 , 1)	-3.3391	0.04%
1S, Masculino, Urbana, Privado	(7 , 0 , 0 , 0)	-2.2481	1.23%
2S, Femenino, Rural, Publico	(8 , 1 , 1 , 1)	-3.6349	0.01%
2S, Femenino, Rural, Privado	(8 , 1 , 1 , 0)	-2.5439	0.55%
2S, Femenino, Urbana, Publico	(8 , 1 , 0 , 1)	-2.7284	0.32%
2S, Femenino, Urbana, Privado	(8 , 1 , 0 , 0)	-1.6374	5.08%
2S, Masculino, Rural, Publico	(8 , 0 , 1 , 1)	-3.8331	0.01%
2S, Masculino, Rural, Privado	(8 , 0 , 1 , 0)	-2.7421	0.31%
2S, Masculino, Urbana, Publico	(8 , 0 , 0 , 1)	-2.9266	0.17%
2S, Masculino, Urbana, Privado	(8 , 0 , 0 , 0)	-1.8356	3.32%
3S, Femenino, Rural, Publico	(9 , 1 , 1 , 1)	-3.2224	0.06%
3S, Femenino, Rural, Privado	(9 , 1 , 1 , 0)	-2.1314	1.65%
3S, Femenino, Urbana, Publico	(9 , 1 , 0 , 1)	-2.3159	1.03%
3S, Femenino, Urbana, Privado	(9 , 1 , 0 , 0)	-1.2249	11.03%
3S, Masculino, Rural, Publico	(9 , 0 , 1 , 1)	-3.4206	0.03%
3S, Masculino, Rural, Privado	(9 , 0 , 1 , 0)	-2.3296	0.99%
3S, Masculino, Urbana, Publico	(9 , 0 , 0 , 1)	-2.5141	0.60%
3S, Masculino, Urbana, Privado	(9 , 0 , 0 , 0)	-1.4231	7.74%
4S, Femenino, Rural, Publico	(10 , 1 , 1 , 1)	-2.8099	0.25%
4S, Femenino, Rural, Privado	(10 , 1 , 1 , 0)	-1.7189	4.28%
4S, Femenino, Urbana, Publico	(10 , 1 , 0 , 1)	-1.9034	2.85%
4S, Femenino, Urbana, Privado	(10 , 1 , 0 , 0)	-0.8124	20.83%
4S, Masculino, Rural, Publico	(10 , 0 , 1 , 1)	-3.0081	0.13%
4S, Masculino, Rural, Privado	(10 , 0 , 1 , 0)	-1.9171	2.76%
4S, Masculino, Urbana, Publico	(10 , 0 , 0 , 1)	-2.1016	1.78%
4S, Masculino, Urbana, Privado	(10 , 0 , 0 , 0)	-1.0106	15.61%
5S, Femenino, Rural, Publico	(11 , 1 , 1 , 1)	-2.3974	0.83%
5S, Femenino, Rural, Privado	(11 , 1 , 1 , 0)	-1.3064	9.57%
5S, Femenino, Urbana, Publico	(11 , 1 , 0 , 1)	-1.4909	6.80%
5S, Femenino, Urbana, Privado	(11 , 1 , 0 , 0)	-0.3999	34.46%
5S, Masculino, Rural, Publico	(11 , 0 , 1 , 1)	-2.5956	0.47%
5S, Masculino, Rural, Privado	(11 , 0 , 1 , 0)	-1.5046	6.62%
5S, Masculino, Urbana, Publico	(11 , 0 , 0 , 1)	-1.6891	4.56%
5S, Masculino, Urbana, Privado	(11 , 0 , 0 , 0)	-0.5981	27.49%

Estas probabilidades fueron calculadas por el autor de la presente tesis

El siguiente cuadro muestra la estimación de la varianza en el segundo nivel, y como se puede apreciar el $p_valor = 0.000 < 0.05$, lo que indicaría que la probabilidad de obtener puntajes mayores o iguales a los de la OECD difieren de colegio en colegio.

Estimated level 2 variances and covariances				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept/intcept	0.4848	0.0813	5.9626	0.0000

En la siguiente tabla se muestra el coeficiente intraclase del modelo, el cual resulta ser 0.326, esto indica que el 32.6% de la variabilidad es explicada por las unidades del segundo nivel.

Calculation of the intracluster correlation	

residual variance = 1 (assumed)	
cluster variance =	0.4848
intracluster correlation =	$0.4848 / (0.4848 + 1.000) = 0.326$

MODELO FINAL PROBIT

```
o=====o
| Modelo Final |
o=====o

Model and Data Descriptions

Sampling Distribution              = Bernoulli
Link Function                     = Probit
PROB(Success)= PHI(ETA)
PHI denotes the CDF of the standard Normal distribution

Level-1 Weight Variable          = Peso
Number of Level-2 Units          = 240
Number of Level-1 Units          = 5985
```

Como podemos apreciar de esta salida veremos que la distribución empleada fue la de Bernoulli con función de enlace Probit, además nos indica cómo se calcularía la probabilidad de éxito. Se consideró los pesos muestrales de los estudiantes, la data cuenta con 5985 estudiantes y 240 colegios.

Sin considerar efectos aleatorios, el ajuste del modelo es de la siguiente forma:

```
o=====o
| Results for the model without any random effects |
o=====o

Goodness of fit statistics

Statistic              Value              DF              Ratio
-----
Likelihood Ratio Chi-square      49552.0212      5974      8.2946
Pearson Chi-square      93202584.1748      5974      15601.3700
```

Considerando al intercepto como efecto aleatorio en el segundo nivel, tenemos que las medidas estadísticas para evaluar la adecuación del modelo son Desvianza con 2452.71, criterio de información de Akaike (AIC) con 2474.71 y el criterio bayesiano de schwarz (BIC) 2548.38 En particular si se desea verificar la

capacidad predictiva del modelo (bondad de ajuste) nuestro estadístico es la desviación $D = 2452.71$ la cual es inferior al punto crítico Chi cuadrado al 95% de nivel de confianza con 5973 grados de libertad $X^2_{(5973)} = 6153.91$, con lo cual no se rechaza la hipótesis nula, es decir el modelo ajustado si es significativo.

<pre> ===== Optimization Method: Adaptive Quadrature ===== </pre>	
Number of quadrature points =	10
Number of free parameters =	11
Number of iterations used =	3
-2lnL (deviance statistic) =	2452.71490
Akaike Information Criterion	2474.71490
Schwarz Criterion	2548.38203

La función de enlace probit será:

$$Prob(XPCL50 = 1) = \Phi^{-1}(\eta_{ij})$$

Modelo expresados en dos niveles serán:

Nivel 1:

$$\eta_{ij} = \beta_{0i} + \beta_{1i} \times XGrado_{ij} + \beta_{2i} \times XSexo_{ij} + \beta_{3i} \times XMetasum_{ij} + \beta_{4i} \times XUndrem_{ij} + \beta_{5i} \times XESCS_{ij} + \varepsilon_{ij}$$

Nivel 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} \times ZZona_j + \gamma_{02} \times ZTipo_j + \gamma_{03} \times ZPropqua_j + \gamma_{04} \times ZSCMated_j + \mu_{0j}$$

$$\beta_{ki} = \gamma_{k0}, \quad k = 1, 2, \dots, 5$$

Otra forma de expresar es:

$$\eta_{ij} = \gamma_{00} + \gamma_{10} \times XGrado_{ij} + \gamma_{20} \times XSexo_{ij} + \gamma_{30} \times XMetasum_{ij} + \gamma_{40} \times XUndrem_{ij} + \gamma_{50} \times XESCS_{ij} + \gamma_{01} \times ZZona_j + \gamma_{02} \times ZTipo_j + \gamma_{03} \times ZPropqua_j + \gamma_{04} \times ZSCMated_j + \mu_{0j} + \varepsilon_{ij}$$

Los coeficientes estimados, considerando el efecto aleatorio en el intercepto es:

Estimated regression weights				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept	-4.4800	0.5043	-8.8836	0.0000
Xgrado	0.3176	0.0466	6.8191	0.0000
XSexo1	0.1652	0.0648	2.5490	0.0108
XMETASUM	0.3770	0.0412	9.1558	0.0000
XUNDREM	0.2393	0.0346	6.9180	0.0000
XESCS	0.2649	0.0345	7.6736	0.0000
ZZONAl	-0.2902	0.2131	-1.3621	0.1732
ZTC1	-0.3585	0.1145	-3.1302	0.0017
ZPROPQUA	0.6610	0.1807	3.6579	0.0003
ZSCMATED	0.2028	0.0465	4.3644	0.0000

Como se aprecia del cuadro la mayor parte de los p_valores de los coeficientes estimados son significativos ($p_valor < 0.05$). El intercepto resulta ser -4.48, es el valor promedio del predictor lineal cuya función de enlace es el probit. En la variable Grado su coeficiente estimado fue de 0.3176, en donde este valor variara debido al grado en que se encuentre cada alumno. En la variable Sexo su coeficiente estimado fue de 0.1652, y será aplicado cuando el sexo sea femenino. En la variable METASUM su coeficiente estimado fue de 0.377, en donde este valor variara debido al puntaje que obtenga cada alumno, es decir por cada unidad que se incremente en la Metacognición en Lectura (Resumir) se estima que el predictor lineal aumentará en promedio 0.377 unidades. En la variable UNDREM su coeficiente estimado fue de 0.2393, en donde este valor variara debido al puntaje que obtenga cada alumno, es decir por cada unidad en que se incremente en la Metacognición en Lectura (comprender y recordar) se estima que el predictor lineal aumentará en promedio en 0.2393 unidades. En la variable ESCS su coeficiente estimado fue de 0.2649, en donde este valor variara debido al puntaje que obtenga cada alumno, es decir por cada unidad que se incremente en el Índice de estatus económico, social y cultural se estima que el predictor lineal aumentara en promedio 0.2649 unidades. En la variable Zona su coeficiente estimado fue de - 0.2902, y será aplicado cuando la zona del colegio sea rural. En

la variable Tipo de colegio su coeficiente estimado fue de -0.3585, y será aplicado cuando el colegio sea público. En la variable PROPQUA su coeficiente estimado fue de 0.6610, es decir por cada unidad que se incremente sobre la proporción de docentes calificados en cada colegio se estima que el predictor lineal se incrementara en promedio en 0.6610 unidades, de manera similar para la variable SCMATED calidad de recursos educativos, donde su coeficiente estimado fue de 0.2028.

Entonces el modelo estimado será:

$$\begin{aligned}\hat{\eta}_{ij} = & -4.48 + 0.3176 \times XGrado_{ij} + 0.1652 \times XSexo_{ij} + 0.377 \times XMetasum_{ij} \\ & + 0.2393 \times XUndrem_{ij} + 0.2649 \times XESCS_{ij} - 0.2902 \times ZZona_j \\ & - 0.3585 \times ZTipo_j + 0.6610 \times ZPropqua_j + 0.2028 \times ZSCMated_j\end{aligned}$$

Las probabilidades asociadas a obtener o superar el promedio de la OECD en comprensión lectora, para un estudiante de un determinado grado, sexo, que haya obtenido un puntaje promedio en METASUM, UNDREM y en ESCS, y en el segundo nivel, el tipo de colegio, la zona en la que se ubique el colegio, en particular con un 50% de docentes calificados (*para el cálculo de la probabilidad se tomó en cuenta el valor 0.5 por el coeficiente 0.6610*) y que haya obtenido un puntaje promedio en la calidad de recursos educativos del colegio (*por estar la variable estandarizada, el valor promedio se ubicaría en cero, por tanto para el cálculo de la probabilidad el coeficiente no afectaría, es por ello que se muestran solo las columnas para algunas variables*), son presentadas en la siguiente tabla:

**Cuadro N°4.4: PROBABILIDADES CORRESPONDIENTES AL
MODELO PROBIT FINAL**

Grupo	Código	η	Probabilidad (η)
1S, Femenino, Rural, Publico	(7 , 1 , 1 , 1)	-2.4098	0.80%
1S, Femenino, Rural, Privado	(7 , 1 , 1 , 0)	-2.0513	2.01%
1S, Femenino, Urbana, Publico	(7 , 1 , 0 , 1)	-2.1196	1.70%
1S, Femenino, Urbana, Privado	(7 , 1 , 0 , 0)	-1.7611	3.91%
1S, Masculino, Rural, Publico	(7 , 0 , 1 , 1)	-2.575	0.50%
1S, Masculino, Rural, Privado	(7 , 0 , 1 , 0)	-2.2165	1.33%
1S, Masculino, Urbana, Publico	(7 , 0 , 0 , 1)	-2.2848	1.12%
1S, Masculino, Urbana, Privado	(7 , 0 , 0 , 0)	-1.9263	2.70%
2S, Femenino, Rural, Publico	(8 , 1 , 1 , 1)	-2.0922	1.82%
2S, Femenino, Rural, Privado	(8 , 1 , 1 , 0)	-1.7337	4.15%
2S, Femenino, Urbana, Publico	(8 , 1 , 0 , 1)	-1.802	3.58%
2S, Femenino, Urbana, Privado	(8 , 1 , 0 , 0)	-1.4435	7.44%
2S, Masculino, Rural, Publico	(8 , 0 , 1 , 1)	-2.2574	1.20%
2S, Masculino, Rural, Privado	(8 , 0 , 1 , 0)	-1.8989	2.88%
2S, Masculino, Urbana, Publico	(8 , 0 , 0 , 1)	-1.9672	2.46%
2S, Masculino, Urbana, Privado	(8 , 0 , 0 , 0)	-1.6087	5.38%
3S, Femenino, Rural, Publico	(9 , 1 , 1 , 1)	-1.7746	3.80%
3S, Femenino, Rural, Privado	(9 , 1 , 1 , 0)	-1.4161	7.84%
3S, Femenino, Urbana, Publico	(9 , 1 , 0 , 1)	-1.4844	6.89%
3S, Femenino, Urbana, Privado	(9 , 1 , 0 , 0)	-1.1259	13.01%
3S, Masculino, Rural, Publico	(9 , 0 , 1 , 1)	-1.9398	2.62%
3S, Masculino, Rural, Privado	(9 , 0 , 1 , 0)	-1.5813	5.69%
3S, Masculino, Urbana, Publico	(9 , 0 , 0 , 1)	-1.6496	4.95%
3S, Masculino, Urbana, Privado	(9 , 0 , 0 , 0)	-1.2911	9.83%
4S, Femenino, Rural, Publico	(10 , 1 , 1 , 1)	-1.457	7.26%
4S, Femenino, Rural, Privado	(10 , 1 , 1 , 0)	-1.0985	13.60%
4S, Femenino, Urbana, Publico	(10 , 1 , 0 , 1)	-1.1668	12.16%
4S, Femenino, Urbana, Privado	(10 , 1 , 0 , 0)	-0.8083	20.95%
4S, Masculino, Rural, Publico	(10 , 0 , 1 , 1)	-1.6222	5.24%
4S, Masculino, Rural, Privado	(10 , 0 , 1 , 0)	-1.2637	10.32%
4S, Masculino, Urbana, Publico	(10 , 0 , 0 , 1)	-1.332	9.14%
4S, Masculino, Urbana, Privado	(10 , 0 , 0 , 0)	-0.9735	16.52%
5S, Femenino, Rural, Publico	(11 , 1 , 1 , 1)	-1.1394	12.73%
5S, Femenino, Rural, Privado	(11 , 1 , 1 , 0)	-0.7809	21.74%
5S, Femenino, Urbana, Publico	(11 , 1 , 0 , 1)	-0.8492	19.79%
5S, Femenino, Urbana, Privado	(11 , 1 , 0 , 0)	-0.4907	31.18%
5S, Masculino, Rural, Publico	(11 , 0 , 1 , 1)	-1.3046	9.60%
5S, Masculino, Rural, Privado	(11 , 0 , 1 , 0)	-0.9461	17.20%
5S, Masculino, Urbana, Publico	(11 , 0 , 0 , 1)	-1.0144	15.52%
5S, Masculino, Urbana, Privado	(11 , 0 , 0 , 0)	-0.6559	25.59%

Estas probabilidades fueron calculadas por el autor de la presente tesis

De la tabla anterior dado que las variables METASUM, UNDREM, ESCS y SCMATED se encuentran estandarizadas, entonces sus promedios son el valor cero, para lo cual se hizo el cálculo de las probabilidades.

El siguiente cuadro muestra la estimación de la varianza en el segundo nivel, y como se puede apreciar el $p_valor = 0.000 < 0.05$, lo que indicaría que la probabilidad de obtener puntajes mayores o iguales a los de la OECD difieren de colegio en colegio.

Estimated level 2 variances and covariances				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept/intcept	0.1882	0.0411	4.5751	0.0000

En la siguiente tabla se muestra el coeficiente intraclase del modelo, el cual resulta ser 0.158, esto indica que el 15.8% de la variabilidad es explicada por las unidades del segundo nivel. (ver página 24[12]).

Calculation of the intracluster correlation	

residual variance = 1 (assumed)	
cluster variance = 0.1882	
intracluster correlation =	$0.1882 / (0.1882 + 1.000) = 0.158$

**Cuadro N°4.5: RESUMEN DE LOS MODELOS DE ANALISIS MULTINIVEL
CON ENLACE LOGIT Y ENLACE PROBIT**

<i>FUNCIÓN</i>		FUNCIÓN DE ENLACE LOGIT			FUNCIÓN DE ENLACE PROBIT		
NIVEL	VARIABLES	MODELO 1	MODELO 2	MODELO 3	MODELO 1	MODELO 2	MODELO 3
INTERCEPTO		-9.5714	-7.5953	-8.3819	-5.1356	-4.0300	-4.4800
MICRO	XGRADO	0.7769	0.5978	0.6015	0.4125	0.3129	0.3176
	XSEXO	0.3731	0.2863	0.2977	0.1982	0.1555	0.1652
	XMETASUM	---	0.7190	0.7010	---	0.3850	0.3770
	XUNDREM	---	0.4336	0.4318	---	0.2399	0.2393
	XESCS	---	0.5174	0.4728	---	0.2878	0.2649
	XDISCLIM	---	0.0638	---	---	0.0439	---
MACRO	ZZONA	-1.8477	-1.0065	-0.6000	-0.9065	-0.5174	-0.2902
	ZTC	-2.0648	-1.2827	-0.6578	-1.0910	-0.6913	-0.3585
	ZPROPQUA	---	---	1.2366	---	---	0.6610
	ZSCMATED	---	---	0.3871	---	---	0.2028

**Cuadro N°4.6: ESTADÍSTICOS DE LOS MODELOS DE ANALISIS MULTINIVEL
CON ENLACE LOGIT Y ENLACE PROBIT**

ESTADISTICOS	FUNCIÓN DE ENLACE LOGIT			FUNCIÓN DE ENLACE PROBIT		
	MODELO 1	MODELO 2	MODELO 3	MODELO 1	MODELO 2	MODELO 3
N° parámetros	6	10	11	6	10	11
varianza estimada del nivel 2	1.74453	0.89359	0.65341	0.48475	0.25644	0.18816
DESVIANZA (D)	2802.29892	2504.78632	2469.17025	2798.14251	2488.39419	2452.71490
Chi cuadrado (punto. Critico)	6158.98388	6154.92369	6153.90864	6158.98388	6154.92369	6153.90864
AKAIKE (AIC)	2814.29892	2524.78632	2491.17025	2810.14251	2508.39419	2474.71490
SCHWARZ (BIC)	2854.48099	2591.75644	2564.83738	2850.32457	2575.36430	2548.38203
AKAIKE corr. (AICC)	2814.31297	2524.82315	2491.21445	2810.15656	2508.43102	2474.75910
AKAIKE cons. (CAIC)	2860.48099	2601.75644	2575.83738	2856.32458	2585.36431	2559.38203
COEFICIENTE INTRACLASE	34.7%	21.4%	16.6%	32.6%	20.4%	15.8%

Los estadísticos Akaike corregido y Akaike consistente han sido calculados por el autor de la presente tesis.

CONCLUSIONES

A medida que se incrementó el número de variables en cada uno de los modelos de análisis multinivel cuya función de enlace es Logit, la varianza estimada del segundo nivel, como la desvianza tiende a reducirse, este mismo comportamiento ocurrió en los modelos propuestos del análisis multinivel con función de enlace Probit. *(Esa misma tendencia se dio para las medidas de calidad del ajuste global en cada uno de los modelos).*

Los coeficientes de correlación intraclase obtenidos en cada uno de los modelos de análisis multinivel Logit se han ido reduciendo, de la siguiente manera: 34.7%, 21.4% y 16.6%, específicamente en el modelo 1, modelo 2 y modelo 3 respectivamente, sin embargo esta disminución se ha dado por encima del 15%. De la misma manera, los coeficientes de correlación intraclase obtenidos en cada uno de los modelos análisis multinivel Probit se han ido reduciendo, de la siguiente manera: 32.6%, 20.4% y 15.8% específicamente en el modelo 1, modelo 2 y modelo 3 respectivamente. Logrando también superar el 15% en estos modelos propuestos.

Considerando únicamente los modelos propuestos del Análisis Multinivel con función Logit *(específicamente en los modelos 1, 2 y 3)* y comparando su capacidad predictiva en cada uno ellos, la desvianza es menor que su respectivo punto crítico en todos los casos, es decir estos modelos si tienen un buen ajuste. Esto mismo ocurre en los modelos propuestos del Análisis Multinivel con función de modelos Probit *(específicamente en los modelos 1, 2 y 3)*

Al comparar los tres modelos Logit en base a las medidas de ajuste global (*en los Criterios de Información de Akaike, Akaike corregido, Akaike consistente, y el criterio de información Bayesiano*) para cada uno de los modelos, se sabe que mientras menores sean estas medidas en un determinado modelo, este será aquel modelo que tiene un mejor ajuste global. En base a este criterio el modelo 3 del Logit es el que mejor ajuste global tiene. Al hacer el análisis para los tres modelos Probit se concluye que también el modelo 3 del Probit es aquel que tiene un mejor ajuste global.

Si bien es cierto que ambos modelos se gestan bajo diferentes objetivos en ambos casos es posible calcular las probabilidades de éxito para cada caso, por tanto es posible realizar clasificaciones en función de las mismas.

Se debe tener un especial cuidado al comparar los modelos Logit y Probit si en especial se tratan de diferencias, debido a que ya existe una diferencia principal. Y es que utilizan funciones de enlace que matemáticamente son diferentes (*Logaritmo neperiano de los odds, y la inversa de la distribución normal estándar acumulada respectivamente.*), debido a ello no se debe esperar que las estimaciones coincidan. Aun así de ser necesario una revisión de sus indicadores entre ellos, en cuanto a las medidas de ajuste global del primer modelo Logit con el primer modelo Probit, es necesario que en ambos casos se tenga el mismo tamaño de muestra, el mismo número de parámetros a estimar, tomando en cuenta las mismas variables independientes y la variable dependiente el cuál se tomó en cuenta en la presente tesis. Tomando en cuenta estos criterios, los Coeficientes de Correlación Intraclase difieren entre ellos, pero esta diferencia no es muy alta entre ellos. Y en particular los coeficientes de correlación son ligeramente menores en los modelos Probit que los modelos obtenidos con el modelo Logit.

En los modelos de análisis multinivel cuya función de enlace es Logit, se pueden obtener los odds ratio OR, que vienen a ser la posibilidad de que ocurra un evento “éxito” dado ciertas características que se dan en la variables

dependientes, en particular estos resultados no se obtienen bajo el modelo Probit.

Cuando se observa las medidas de calidad el modelo 1 del Logit con las del modelo Probit estas son diferentes, pero a la vez cercanas, siendo menores en todos los casos en el modelo Probit, esto mismo ocurre para los modelos 2 y 3 del modelo Logit y Probit respectivamente.

Se calcularon las probabilidades de éxito, para diferentes casos, bajo los mismos criterios. Estas probabilidades difieren entre los modelos Logit y Probit, a pesar que existen diferencias entre las probabilidades calculadas las tendencias no cambian es decir si se aprecia claramente en el modelo Logit que un caso particular tiene mayor probabilidad de éxito que otro caso particular, esta ventaja se muestra también en el modelo Probit.

Las características que se consideraron en el primer modelo para poder predecir las probabilidades de éxito (*que un estudiante supere el promedio OCDE, en comprensión de lectura*), fue el grado del estudiante, sexo, y también la ubicación del colegio (zona: rural y urbana), como también el tipo de colegio respecto a si es público o privado encontrándose que estas características son significativas en el estudio del rendimiento académico, y en especial se atribuye que el 34.7% y 32.6% del total de la varianza explicada es debido a los colegios, en los modelos de análisis multinivel con las funciones de enlace Logit y Probit respectivamente. **Es decir los colegios tienen un efecto muy importante sobre el rendimiento de los estudiantes.**

El porcentaje de varianza explicada debido a los colegios es importante, en el estudio del rendimiento académico en cuanto a comprensión lectora, Sin embargo aun así este efecto disminuye cuando se agregaron características propias del estudiante como la Metacognición o cuando se incrementó el Estatus económico, social y cultural, entre otras, es decir características que están relacionadas con su entorno familiar del propio estudiante. Si bien es cierto que el colegio cumple un rol muy importante, y no se le puede restar importancia, de la misma manera no se puede atribuir toda la responsabilidad,

debido a que hay aspectos inherentes propios del ser, en este caso del estudiante y de la familia, en los cuales se debe buscar la fórmula para incidir en ellos para poder motivar a que nuestros estudiantes se inclinen hacia el desarrollo personal, en función de la adquisición de habilidades, y en particular en la comprensión de lectura.

Al analizar las probabilidades en los diferentes modelos, ya sea con función de enlace Logit o Probit, la probabilidad que un estudiante supere el promedio de la OCDE en comprensión de lectura, dadas las diversas características de los estudiantes, se obtuvo probabilidades superiores en la mayoría de los casos cuando los estudiantes provienen de un colegio privado en comparación con los estudiantes de colegios públicos. Es importante que se analice por qué se da esta diferencia en cuanto a las probabilidades de éxito, entre los colegios públicos y privados, o bien preguntarnos a partir de qué momento se dio esta diferencia en el país y a que se debió, con la intención de tomar las medidas adecuadas para que nuestros estudiantes tengan las mismas oportunidades de éxito.

Al haber realizado el análisis de manera exclusiva al contexto de nuestro país nos permite conocer cuáles son las fortalezas que tiene un determinado estudiante peruano en comparación de otro estudiante, y consigo poder cuantificar estas ventajas. Para así hacerla extensiva a todos nuestros estudiantes. Es decir, es importante aumentar el número de colegios, pero esta importancia aumentaría si va acompañada de una adecuada implementación de materiales educativos de calidad y contar con profesores calificados. Por otro lado si homogenizamos la gestión en los colegios este efecto sería aun mayor debido a que se pudo comprobar el efecto positivo que tienen los estudiantes que provienen de colegios privados a pesar que no todos colegios privados tienen los mismos recursos económicos. A partir de ello analizar su fortalezas desde su gestión entre otros factores sería muy interesante.

LIMITACIONES Y **RECOMENDACIONES**

Al utilizar programas computacionales (software) orientado a estudiantes, se tiene restricciones respecto al número de variables, número de niveles o bien el número de casos. *(De ser necesario trabajar con un mayor número de variables se recomendaría tomar en cuenta los aspectos técnicos, como la adquisición de un software).*

Si se desea realizar análisis en el cual se sospecha que existen interacciones entre las variables se debe tomar en cuenta que cada interacción es considerada como una nueva variable por ende se recomendaría considerar este aspecto en caso de tener limitaciones con número de variables.

En la presente tesis se ha trabajado dos casos particulares de los modelos no lineales del análisis multinivel (Logit y Probit), se podría tomar en cuenta para quienes desean investigar algunas variantes de estos mismos modelos bajo la estructura de los Modelos Multinivel, como también para otros modelos no lineales.

Los datos utilizados para el análisis del rendimiento académico en comprensión lectora pertenecen a un determinado año en el Perú (2009), de haber interés en comparar los resultados obtenidos en comprensión lectora (ya sea en ciencias o matemáticas), entonces sería posible estudiar los cambios y/o el comportamiento de los resultados obtenidos por nuestros estudiantes a través del tiempo esto sería de una gran utilidad para conocer nuestro progreso en nuestro contexto peruano.

BIBLIOGRAFÍA

- [1]. Montero, E., Villalobos, J., Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: Un análisis multinivel. Publicado por RELIEVE (Revista Electrónica de Investigación y Evaluación Educativa), v. 13, n. 2, p. 215-234. Obtenido el 18/10/2013 desde www.uv.es/RELIEVE/v13n2/RELIEVEv13n2_5.htm

- [2]. Álvarez, D. y otros. (2006 - 2007) Metodología y logística de campo de un estudio Multinivel sobre la influencia en España de las características medioambientales en la salud mental de población autóctona y ecuatoriana inmigrante. Publicado por el Instituto de Salud Carlos III, Centro Nacional de Epidemiología, Madrid, España. Obtenido el 14/02/2012 desde <http://www.scielosp.org/pdf/resp/v83n4/colaboracion1.pdf>.

- [3]. Cervini, R. (2006). Artículo: Trabajo infantil y progreso de aprendizaje en la educación básica. Publicado por la Revista Latinoamericana de Estudios Educativos; Jul 2006; 36, 3; PRISMA (Publicaciones y Revistas Sociales y Humanísticas) pg. 183. Obtenido el 18/10/2013 desde <http://www.redalyc.org/pdf/270/27036409.pdf>

- [4]. Calero, J., Escardíbul, J., Waisgrais, S., Mediavilla, M. (2003). Desigualdades socioeconómicas en el sistema educativo español. Publicado por Centro de Investigación y Documentación Educativa (Ministerio de Educación y Ciencia) y la Fundación Bosch i Gimpera (Universidad de Barcelona), España. Obtenido el 22/02/2013 desde http://www.oei.es/pdf2/desigualdades_socioeduc_espana.pdf.

- [5]. De Onis, M., Frongillo, E., Blossner, M. (2001). Análisis de la evolución del

nivel de malnutrición infantil desde 1980. Publicado por la Organización Mundial de la Salud, recopilación de artículos N°4, 2001. Obtenido el 22/02/2013 desde http://www.who.int/nutgrowthdb/publications/en/bu0688_esp.pdf.

- [6]. Benavides, M., Caro, D., Espinosa, G., Miranda, L., Tam, M., Zacharías, D., Zambrano, G. (2004). Factores asociados al rendimiento estudiantil. Publicado por el Ministerio de Educación, Perú. Obtenido el 22/02/2013 desde http://www2.minedu.gob.pe/umc/admin/images/documentos/archivo_9.pdf

- [7]. Andréu, J. (2011). El Análisis Multinivel: Una revisión actualizada en el ámbito sociológico. Publicado por el Departamento de Sociología. Universidad de Granada. Metodología de Encuestas, Volumen 13, 2011, 161-176, ISSN: 1575-7803. Obtenido el 22/02/2013 desde <http://casus.usal.es/pkp/index.php/MdE/article/view/1017/958>

- [8]. Pérez, V. (2013) Los modelos multinivel en el análisis de factores de riesgo de sibilancias recurrentes en lactantes. Enfoques frecuentista y bayesiano. Publicado por la Universidad de Murcia, España. Obtenido el 28/10/13 desde <http://www.tdx.cat/bitstream/handle/10803/109213/TVPF.pdf?sequence=1>.

- [9]. Jöreskog, K., Söborn, D. (2013) Multilevel Generalized Linear Modeling Guide. Publicado dentro del software LISREL versión 9.10 para estudiantes.

- [10]. Snijders, T., and Bosker, R., Multilevel analysis. An introduction to basic and advanced multilevel modeling. Sage Publications, London.

- [11]. Goldstein, H. (1999) Multilevel Statistical Models. Publicado por el London: Institute of Education. Obtenido el 22/02/2013 desde <http://www.arnoldpublishers.com/support/goldstein.htm>

- [12]. Micceri, T. (2007) The Advantages of Using Multilevel Modeling to Address Institutional Research Questions. Publicado en Florida Association for Institutional Research Annual Forum, Cocoa Beach, FL, Feb 7-9, 2007. Obtenido el 28/10/13 desde <http://www.florida-air.org/ted07.pdf>
- [13]. Lemonte, A., Vanegas, L. (2005) Una comparación entre la inferencia basada en la estadísticas de wald y razón de verosimilitud en los modelos logit y probit vía Monte Carlo. Publicado por la Revista Colombiana de Estadística Volumen 28 N°1. pp. 77 a 96, Junio 2005. Obtenido el 22/02/2013 desde <http://www.revistas.unal.edu.co/index.php/estad/article/viewFile/28817/29075>
- [14]. Sheldon, R. (1999) Simulación. PRENTICE HALL 2da edición, México.
- [15]. Hosmer, D., Lemeshow, S. (2000), Applied Logistic Regression. Wiley & Sons, New York.
- [16]. Pardo, A., Ruiz, M., San Martín, R. (2007) Como ajustar e interpretar modelos multinivel con SPSS. Publicado por la Universidad Autónoma de Madrid. Obtenido el 22/02/2013 desde <http://www.psicothema.com/psicothema.asp?id=3365>
- [17]. Luppi, I., Boggio, G. (2006). Los modelos de niveles múltiples: una estrategia analítica para el estudio de los problemas de salud de la población. Publicado en la Revista Brasileira de Epidemiologia vol9 n°1 Sao Paulo. Obtenido el 20/10/2013 desde <http://bvsaud.org/portal/resource/pt/lil-430471>.
- [18]. McCulloch C, Searle S. Generalized, linear and mixed models. New York: John Wiley & Sons; 2002.
- [19]. Demidenko, E. Mixed Models. Theory and Applications. New Jersey: John

Wiley & Sons; 2004.

- [20]. Richard Buxton. 2008. Statistics: Multilevel modeling. Publicado por Mathematics Learning, Support Centre. Obtenido el 28/10/2013 desde <http://www.statstutor.ac.uk/resources/uploaded/multilevelmodelling.pdf>
- [21]. Aparicio, A., Morera, M. (2007) La conveniencia del análisis multinivel para la investigación en salud: una aplicación para Costa Rica. Publicado en Población y Salud en Mesoamérica, vol. 4 no. 2, de la Universidad de Costa Rica. Obtenido el 28/10/13 desde <http://www.redalyc.org/articulo.oa?id=44640206>
- [22]. De la Cruz, F. (2008) Modelos Multinivel. Sección Epidemiología del Instituto de Medicina Tropical “Daniel A. Carrión”, UNMSM, Perú. Publicado en la Revista Peruana de Epidemiología Vol. 12 N°3. Obtenido el 22/02/2012 desde http://rpe.epiredperu.net/rpe_ediciones/v12_n03_2008/AR1.pdf#22
- [23]. Llano, L., Mosquera, V. (2006) El modelo LOGIT una alternativa para medir probabilidad de permanencia estudiantil. Publicado por la Facultad de Administración de la Universidad Nacional de Colombia. Obtenido el 28/10/13 desde <http://www.bdigital.unal.edu.co/1038/1/laurarosallanodiaz.2006.pdf>.
- [24]. Rasbash Jon, Steele Fiona, Browne William y Prosser Bob (2004). A User's Guide to MLwiN Version 2.0. Centre for Multilevel Modelling Institute of Education University of London, Printed November 2004. Obtenido el 29/10/13 desde <http://www.mlwin.com/download/userman20.pdf>.
- [25]. Rojas, L. (2013) Predicción de la dificultad de la prueba de habilidades cuantitativas de la Universidad de Costa Rica. Publicado por la Revista digital Matemática, Educación e Internet Vol. 13 No 1. Obtenido el 28/10/13 desde <http://www.tec->

digital.itcr.ac.cr/revistamatematica/ARTICULOS_V13_N1_2012/RevistaDigital_Rojas_V13_n1_2012/RevistaDigital_Rojas_V13_n1_2012.pdf

- [26]. Preisser, J. (2004) The first autor replies. Publicado por American Journal of Epidemiology, Volume 160. Obtenido el 29/10/13 desde <http://aje.oxfordjournals.org/content/160/5/506.full#ref-1>.
- [27]. Gujarati, D. (2006) Principios de Econometria. Publicado por McGrawHill, tercera edición, España.
- [28]. IDRE (2013) SAS Data Analysis Examples Probit Regression. Publicado por University of California, Los Angeles. Obtenido el 28/10/13 desde <http://www.ats.ucla.edu/stat/sas/dae/probit.htm>
- [29]. Ramirez, M. (2005) La migración de retorno en México y la decisión de autoempleo. Publicado por la Universidad de las Américas Puebla. Obtenido el 29/10/13 desde http://catarina.udlap.mx/u_dl_a/tales/documentos/lec/ramirez_r_ma/
- [30]. Enchautegui, M. (2008) Módulo de estudio sobre Modelos Probit y Logit. Publicado por Departamento de Economía de la Universidad de Puerto Rico Recinto de Río Piedras. Revisado el 29/10/13 desde <http://economia.uprrp.edu/notas%20de%20clase%207.pdf>
- [31]. Fox, J. (2010) Notes: Logit and Probit Models. Publicado en York Summer Programme in Data Analysis. Obtenido el 29/10/13 desde <http://socserv.socsci.mcmaster.ca/jfox/Courses/SPIDA/logit-models-notes.pdf>
- [32]. PISA (2009) Informe PISA 2009. Publicado por la OECD, Volumen I. Obtenido el 24/10/12 desde http://www.keepeek.com/Digital-Asset-Management/oecd/education/informe-pisa-2009-lo-que-los-estudiantes-saben-y-pueden-hacer_9789264174900-es
- [33]. Ochoa, S., Aragón, L. (2007) Funcionamiento Meta cognitivo de estudiantes universitarios durante la escritura de reseñas analíticas.

Publicado en la Revista: Universitas Psychologica vol.6 n°3 Bogota.
Obtenido el 23/05/13 desde
[http://www.scielo.org.co/scielo.php?pid=S1657-92672007000300003](http://www.scielo.org.co/scielo.php?pid=S1657-92672007000300003&script=sci_arttext)
&script=sci_arttext.

- [34]. Peronard, M. (2005) La Metacognición como herramienta didáctica.
Publicado en la Revista Signos v.38 n°57 Valparaiso CHILE. Obtenido el
24/09/13 desde [http://www.scielo.cl/scielo.php?pid=S0718-](http://www.scielo.cl/scielo.php?pid=S0718-09342005000100005&script=sci_arttext)
09342005000100005&script=sci_arttext).

ANEXOS

ANEXO 1: Modelo 1 – Logit

```
o-----
--o
|
|           Student Edition of MGLIM Module. (2012)
|
|           Generalized Linear Modeling
|
|           of
|
|           Multilevel (Hierarchical) Data
|
|
|
|           Copyright by Scientific Software International, Inc., 2012
|
|           Scientific Software International, Inc.
|
|           Website:  www.ssicentral.com
|
|           Support:  techsupport@ssicentral.com
|
|
|           DATE OF ANALYSIS: September 13, 2013
|
|           TIME OF ANALYSIS: 21H01:37
|
|           DATA FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
|           LISREL\DataImputada |
|           .LSF
|
|           INPUT FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
|           LISREL\DataImputada |
|           .PRL
|
|
|
o-----
--o

The following lines were read from the MAPGLIM syntax file:
MGLimOptions Converge=0.0001 MaxIter=100 MissingCode=9999
                Method=Quad NQUADPTS=10 ;
Title=Modelo 1;
SY='C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada.LSF';
ID2=SCHOOLID;
Weight=Peso;
Distribution=BER;
Link=LOGIT;
Intercept=Yes;
DepVar=XPCL50;
CoVars=Xgrado XSexo$ ZZONA$ ZTC$;
RANDOM2=intcept;

o=====o
| Modelo 1 |
o=====o
```

Model and Data Descriptions

Sampling Distribution = Bernoulli
 Link Function = Logistic
 PROB(Success)= 1.0/[1.0+EXP(-ETA)]

Level-1 Weight Variable = Peso
 Number of Level-2 Units = 240
 Number of Level-1 Units = 5985
 Number of Level-1 Units per Level-2 Unit =

9	31	16	6	8	35	33	32	4	31	31	10
11	18	32	29	19	29	30	19	31	31	30	31
32	32	35	2	35	29	12	29	10	4	33	32
15	32	11	35	31	9	15	27	30	33	9	33
31	7	32	14	29	32	9	33	31	33	29	32
20	27	8	21	29	3	30	33	26	30	32	32
30	33	32	33	35	24	32	34	31	35	6	34
32	32	16	10	9	13	32	33	32	27	25	35
34	34	11	9	30	33	33	24	32	32	32	3
31	24	33	33	34	8	33	32	9	33	30	28
31	33	30	13	32	31	22	11	27	32	34	33
31	34	33	28	3	9	33	12	30	34	28	33
32	35	9	9	16	11	33	17	21	28	32	34
32	33	27	35	2	8	33	24	28	34	14	9
33	35	27	34	28	29	29	32	31	32	2	32
23	21	30	31	3	12	30	20	33	3	28	3
10	30	33	30	32	32	13	5	8	8	35	8
9	32	32	32	28	2	22	33	33	29	31	35
32	4	30	33	32	34	19	24	31	7	32	8
33	35	33	31	7	10	32	34	35	32	33	34

=====o
 | Descriptive statistics for all the variables in the model |
 o=====o

Variable	Minimum	Maximum	Mean	Standard Deviation
XPCL501	0.0000	1.0000	0.9004	0.2995
XPCL502	0.0000	1.0000	0.0996	0.2995
intcept	1.0000	1.0000	1.0000	0.0000
Xgrado	7.0000	11.0000	9.8251	1.0177
XSexo1	0.0000	1.0000	0.4987	0.5000
ZZONA1	0.0000	1.0000	0.1272	0.3332
ZTC1	0.0000	1.0000	0.7866	0.4097

=====o
 | Results for the model without any random effects |
 o=====o

Goodness of fit statistics

Statistic	Value	DF
Ratio		
---	---	---
Likelihood Ratio Chi-square	33817.5076	5979
5.6560		
Pearson Chi-square	301924.1381	5979
50.4974		

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----

intcept	-9.2638	0.6997	-13.2391	0.0000
Xgrado	0.7844	0.0673	11.6529	0.0000
XSexo1	0.4346	0.0942	4.6132	0.0000
ZZONA1	-1.4753	0.3126	-4.7201	0.0000
ZTC1	-1.7063	0.0937	-18.2028	0.0000

Odds Ratio and 95% Odds Ratio Confidence Intervals

Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
intcept	-9.2638	0.0001	0.0000	0.0004
Xgrado	0.7844	2.1911	1.9202	2.5001
XSexo1	0.4346	1.5444	1.2840	1.8577
ZZONA1	-1.4753	0.2287	0.1239	0.4220
ZTC1	-1.7063	0.1815	0.1511	0.2182

```

=====o
| Optimization Method: Adaptive Quadrature |
=====o

```

```

Number of quadrature points =      10
Number of free parameters =       6
Number of iterations used =       3

```

```

-2lnL (deviance statistic) =      2802.29892
Akaike Information Criterion      2814.29892
Schwarz Criterion                2854.48099

```

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-9.5714	0.8939	-10.7071	0.0000
Xgrado	0.7769	0.0842	9.2299	0.0000
XSexo1	0.3731	0.1177	3.1701	0.0015
ZZONA1	-1.8477	0.4904	-3.7678	0.0002
ZTC1	-2.0648	0.2573	-8.0256	0.0000

Odds Ratio and 95% Odds Ratio Confidence Intervals

Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
intcept	-9.5714	0.0001	0.0000	0.0004
Xgrado	0.7769	2.1746	1.8439	2.5647
XSexo1	0.3731	1.4523	1.1531	1.8291
ZZONA1	-1.8477	0.1576	0.0603	0.4121
ZTC1	-2.0648	0.1268	0.0766	0.2100

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept/intcept	1.7445	0.3043	5.7332	0.0000

Calculation of the intracluster correlation

```

-----
residual variance = pi*pi / 3 (assumed)
cluster variance =      1.7445

```

intracluster correlation = $1.7445 / (1.7445 + (\pi^2/3)) = 0.347$

o=====o
| LISREL Multilevel GLIM used 0.28 seconds CPU |
o=====o

ANEXO 2: Modelo 2 – Logit

```
o-----
--o
|
|           Student Edition of MGLIM Module. (2012)
|
|           Generalized Linear Modeling
|
|           of
|
|           Multilevel (Hierarchical) Data
|
|
|           Copyright by Scientific Software International, Inc., 2012
|
|           Scientific Software International, Inc.
|
|           Website:  www.ssicentral.com
|
|           Support:  techsupport@ssicentral.com
|
|
|           DATE OF ANALYSIS: September 13, 2013
|
|           TIME OF ANALYSIS: 21H19:14
|
|           DATA FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
|           LISREL\DataImputada |
|           .LSF
|
|           INPUT FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
|           LISREL\DataImputada |
|           .PRL
|
|-----o
--o

The following lines were read from the MAPGLIM syntax file:
MGLimOptions Converge=0.0001 MaxIter=100 MissingCode=9999
          Method=Quad NQUADPTS=10 ;
Title=Modelo 2;
SY='C:\Users\user\Desktop\Avances 04-09-13\bd imputada
final\LISREL\DataImputada .LSF';
ID2=SCHOOLID;
Weight=Peso;
Distribution=BER;
Link=LOGIT;
Intercept=Yes;
DepVar=XPCL50;
CoVars=Xgrado XSexo$ XMETASUM XUNDREM XESCS XDISCLIM ZZONA$ ZTC$;
RANDOM2=intcept;

o=====o
| Modelo 2 |
o=====o

          Model and Data Descriptions

          Sampling Distribution          = Bernoulli
          Link Function                 = Logistic
```

PROB(Success)= 1.0/[1.0+EXP(-ETA)]
 Level-1 Weight Variable = Peso
 Number of Level-2 Units = 240
 Number of Level-1 Units = 5985
 Number of Level-1 Units per Level-2 Unit =

9	31	16	6	8	35	33	32	4	31	31	10
11	18	32	29	19	29	30	19	31	31	30	31
32	32	35	2	35	29	12	29	10	4	33	32
15	32	11	35	31	9	15	27	30	33	9	33
31	7	32	14	29	32	9	33	31	33	29	32
20	27	8	21	29	3	30	33	26	30	32	32
30	33	32	33	35	24	32	34	31	35	6	34
32	32	16	10	9	13	32	33	32	27	25	35
34	34	11	9	30	33	33	24	32	32	32	3
31	24	33	33	34	8	33	32	9	33	30	28
31	33	30	13	32	31	22	11	27	32	34	33
31	34	33	28	3	9	33	12	30	34	28	33
32	35	9	9	16	11	33	17	21	28	32	34
32	33	27	35	2	8	33	24	28	34	14	9
33	35	27	34	28	29	29	32	31	32	2	32
23	21	30	31	3	12	30	20	33	3	28	3
10	30	33	30	32	32	13	5	8	8	35	8
9	32	32	32	28	2	22	33	33	29	31	35
32	4	30	33	32	34	19	24	31	7	32	8
33	35	33	31	7	10	32	34	35	32	33	34

o=====o
 | Descriptive statistics for all the variables in the model |
 o=====o

Variable	Minimum	Maximum	Mean	Standard Deviation
-----	-----	-----	----	-----
XPCL501	0.0000	1.0000	0.9004	0.2995
XPCL502	0.0000	1.0000	0.0996	0.2995
intcept	1.0000	1.0000	1.0000	0.0000
Xgrado	7.0000	11.0000	9.8251	1.0177
XSex01	0.0000	1.0000	0.4987	0.5000
XMETASUM	-2.0098	1.3443	-0.3356	0.9033
XUNDREM	-1.8824	1.4152	-0.4322	0.9311
XESCS	-4.8025	1.9933	-1.2962	1.2270
XDISCLIM	-2.8091	1.8379	0.1891	0.7686
ZZONA1	0.0000	1.0000	0.1272	0.3332
ZTC1	0.0000	1.0000	0.7866	0.4097

o=====o
 | Results for the model without any random effects |
 o=====o

Goodness of fit statistics		
Statistic	Value	DF
Ratio	-----	--
---		--
Likelihood Ratio Chi-square	41825.6452	5975
7.0001		
Pearson Chi-square	1098538.5187	5975
183.8558		

Estimated regression weights				
Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept	-6.5414	0.7959	-8.2190	0.0000
Xgrado	0.5205	0.0769	6.7651	0.0000
XSex01	0.3346	0.1034	3.2368	0.0012
XMETASUM	0.7063	0.0723	9.7747	0.0000

XUNDREM	0.4425	0.0602	7.3514	0.0000
XESCS	0.7386	0.0528	13.9877	0.0000
XDISCLIM	0.0577	0.0669	0.8635	0.3879
ZZONA1	-0.6918	0.3246	-2.1313	0.0331
ZTC1	-0.8968	0.1086	-8.2608	0.0000

Odds Ratio and 95% Odds Ratio Confidence Intervals

Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
intcept	-6.5414	0.0014	0.0003	0.0069
Xgrado	0.5205	1.6828	1.4473	1.9567
XSexol	0.3346	1.3973	1.1411	1.7111
XMETASUM	0.7063	2.0265	1.7589	2.3348
XUNDREM	0.4425	1.5566	1.3833	1.7515
XESCS	0.7386	2.0931	1.8873	2.3213
XDISCLIM	0.0577	1.0594	0.9293	1.2078
ZZONA1	-0.6918	0.5007	0.2650	0.9459
ZTC1	-0.8968	0.4079	0.3297	0.5046

o=====o

| Optimization Method: Adaptive Quadrature |

o=====o

Number of quadrature points = 10
Number of free parameters = 10
Number of iterations used = 3
-2lnL (deviance statistic) = 2504.78632
Akaike Information Criterion 2524.78632
Schwarz Criterion 2591.75644

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-7.5953	0.9234	-8.2251	0.0000
Xgrado	0.5978	0.0879	6.8037	0.0000
XSexol	0.2863	0.1220	2.3465	0.0190
XMETASUM	0.7190	0.0778	9.2462	0.0000
XUNDREM	0.4336	0.0645	6.7235	0.0000
XESCS	0.5174	0.0648	7.9833	0.0000
XDISCLIM	0.0638	0.0743	0.8583	0.3907
ZZONA1	-1.0065	0.4398	-2.2888	0.0221
ZTC1	-1.2827	0.2142	-5.9889	0.0000

Odds Ratio and 95% Odds Ratio Confidence Intervals

Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
intcept	-7.5953	0.0005	0.0001	0.0031
Xgrado	0.5978	1.8182	1.5305	2.1599
XSexol	0.2863	1.3315	1.0483	1.6913
XMETASUM	0.7190	2.0524	1.7622	2.3903
XUNDREM	0.4336	1.5428	1.3596	1.7507
XESCS	0.5174	1.6777	1.4776	1.9050
XDISCLIM	0.0638	1.0659	0.9214	1.2330
ZZONA1	-1.0065	0.3655	0.1544	0.8654
ZTC1	-1.2827	0.2773	0.1822	0.4219

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept/intcept	0.8936	0.1888	4.7333	0.0000

Calculation of the intracluster correlation

residual variance = $\pi^2 / 3$ (assumed)

cluster variance = 0.8936

intracluster correlation = $0.8936 / (0.8936 + (\pi^2 / 3)) = 0.214$

o=====o

| LISREL Multilevel GLIM used 0.40 seconds CPU |

o=====o

ANEXO 3: Modelo 3 – Logit

```
o-----
--o
|
|           Student Edition of MGLIM Module. (2012)
|
|           Generalized Linear Modeling
|
|           of
|
|           Multilevel (Hierarchical) Data
|
|
|
|           Copyright by Scientific Software International, Inc., 2012
|
|           Scientific Software International, Inc.
|
|           Website:  www.ssicentral.com
|
|           Support:  techsupport@ssicentral.com
|
|
|           DATE OF ANALYSIS: September 13, 2013
|
|           TIME OF ANALYSIS: 21H21:38
|
|           DATA FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
|           LISREL\DataImputada |
|           .LSF
|
|           INPUT FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
|           LISREL\DataImputada |
|           .PRL
|
|
|
o-----
--o

The following lines were read from the MAPGLIM syntax file:
MGLimOptions Converge=0.0001 MaxIter=100 MissingCode=9999
                Method=Quad NQUADPTS=10 ;
Title=Modelo 3;
SY='C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada.LSF';
ID2=SCHOOLID;
Weight=Peso;
Distribution=BER;
Link=LOGIT;
Intercept=Yes;
DepVar=XPCL50;
CoVars=Xgrado XSexo$ XMETASUM XUNDREM XESCS ZZONA$ ZTC$
        ZPROPQUA ZSCMATED;
RANDOM2=intcept;

o=====o
| Modelo 3 |
o=====o
```

Model and Data Descriptions

Sampling Distribution = Bernoulli
 Link Function = Logistic
 $PROB(Success) = 1.0 / [1.0 + EXP(-ETA)]$

Level-1 Weight Variable = Peso
 Number of Level-2 Units = 240
 Number of Level-1 Units = 5985
 Number of Level-1 Units per Level-2 Unit =

9	31	16	6	8	35	33	32	4	31	31	10
11	18	32	29	19	29	30	19	31	31	30	31
32	32	35	2	35	29	12	29	10	4	33	32
15	32	11	35	31	9	15	27	30	33	9	33
31	7	32	14	29	32	9	33	31	33	29	32
20	27	8	21	29	3	30	33	26	30	32	32
30	33	32	33	35	24	32	34	31	35	6	34
32	32	16	10	9	13	32	33	32	27	25	35
34	34	11	9	30	33	33	24	32	32	32	3
31	24	33	33	34	8	33	32	9	33	30	28
31	33	30	13	32	31	22	11	27	32	34	33
31	34	33	28	3	9	33	12	30	34	28	33
32	35	9	9	16	11	33	17	21	28	32	34
32	33	27	35	2	8	33	24	28	34	14	9
33	35	27	34	28	29	29	32	31	32	2	32
23	21	30	31	3	12	30	20	33	3	28	3
10	30	33	30	32	32	13	5	8	8	35	8
9	32	32	32	28	2	22	33	33	29	31	35
32	4	30	33	32	34	19	24	31	7	32	8
33	35	33	31	7	10	32	34	35	32	33	34

=====O
 | Descriptive statistics for all the variables in the model |
 =====O

Variable	Minimum	Maximum	Mean	Standard Deviation
-----	-----	-----	----	-----
XPCL501	0.0000	1.0000	0.9004	0.2995
XPCL502	0.0000	1.0000	0.0996	0.2995
intcept	1.0000	1.0000	1.0000	0.0000
Xgrado	7.0000	11.0000	9.8251	1.0177
XSexo1	0.0000	1.0000	0.4987	0.5000
XMETASUM	-2.0098	1.3443	-0.3356	0.9033
XUNDREM	-1.8824	1.4152	-0.4322	0.9311
XESCS	-4.8025	1.9933	-1.2962	1.2270
ZZONA1	0.0000	1.0000	0.1272	0.3332
ZTC1	0.0000	1.0000	0.7866	0.4097
ZPROPQUA	0.0000	1.0000	0.4573	0.2812
ZSCMATED	-3.3887	1.9276	-1.1049	1.1895

=====O
 | Results for the model without any random effects |
 =====O

Goodness of fit statistics

Statistic	Value	DF
Ratio	-----	--
---	-----	---
Likelihood Ratio Chi-square	42744.2472	5974
7.1550		
Pearson Chi-square	1305124.3031	5974
218.4674		

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-7.4291	0.8306	-8.9444	0.0000
Xgrado	0.5378	0.0790	6.8109	0.0000
XSexo1	0.3138	0.1043	3.0082	0.0026
XMETASUM	0.6753	0.0730	9.2512	0.0000
XUNDREM	0.4335	0.0607	7.1392	0.0000
XESCS	0.6358	0.0545	11.6601	0.0000
ZZONA1	-0.4150	0.3276	-1.2671	0.2051
ZTC1	-0.4872	0.1219	-3.9954	0.0001
ZPROPQUA	1.0312	0.2100	4.9103	0.0000
ZSCMATED	0.2785	0.0506	5.5066	0.0000

Odds Ratio and 95% Odds Ratio Confidence Intervals

Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
intcept	-7.4291	0.0006	0.0001	0.0030
Xgrado	0.5378	1.7122	1.4667	1.9988
XSexo1	0.3138	1.3686	1.1155	1.6790
XMETASUM	0.6753	1.9646	1.7027	2.2668
XUNDREM	0.4335	1.5427	1.3696	1.7377
XESCS	0.6358	1.8886	1.6972	2.1017
ZZONA1	-0.4150	0.6603	0.3475	1.2548
ZTC1	-0.4872	0.6143	0.4837	0.7802
ZPROPQUA	1.0312	2.8045	1.8582	4.2326
ZSCMATED	0.2785	1.3211	1.1964	1.4588

```

o=====o
| Optimization Method: Adaptive Quadrature |
o=====o

```

```

Number of quadrature points =      10
Number of free parameters =      11
Number of iterations used =        3

```

```

-2lnL (deviance statistic) =      2469.17025
Akaike Information Criterion      2491.17025
Schwarz Criterion                2564.83738

```

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-8.3819	0.9542	-8.7838	0.0000
Xgrado	0.6015	0.0882	6.8204	0.0000
XSexo1	0.2977	0.1205	2.4711	0.0135
XMETASUM	0.7010	0.0776	9.0336	0.0000
XUNDREM	0.4318	0.0642	6.7315	0.0000
XESCS	0.4728	0.0642	7.3586	0.0000
ZZONA1	-0.6000	0.4241	-1.4149	0.1571
ZTC1	-0.6578	0.2150	-3.0589	0.0022
ZPROPQUA	1.2366	0.3448	3.5859	0.0003
ZSCMATED	0.3871	0.0888	4.3580	0.0000

Odds Ratio and 95% Odds Ratio Confidence Intervals

Bounds

Parameter	Estimate	Odds Ratio	Lower	Upper
-----	-----	-----	-----	-----
intcept	-8.3819	0.0002	0.0000	0.0015
Xgrado	0.6015	1.8248	1.5351	2.1691
XSexol	0.2977	1.3468	1.0635	1.7055
XMETASUM	0.7010	2.0158	1.7313	2.3469
XUNDREM	0.4318	1.5401	1.3581	1.7464
XESCS	0.4728	1.6045	1.4146	1.8198
ZZONA1	-0.6000	0.5488	0.2390	1.2601
ZTC1	-0.6578	0.5180	0.3398	0.7895
ZPROPQUA	1.2366	3.4438	1.7519	6.7698
ZSCMATD	0.3871	1.4727	1.2374	1.7528

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept/intcept	0.6534	0.1462	4.4691	0.0000

Calculation of the intracluster correlation

residual variance = $\pi^2 / 3$ (assumed)
cluster variance = 0.6534

intracluster correlation = $0.6534 / (0.6534 + (\pi^2 / 3)) = 0.166$

o=====o
| LISREL Multilevel GLIM used 0.39 seconds CPU |
o=====o

ANEXO 4: Modelo 1 – Probit

```
o-----
--o
|
|           Student Edition of MGLIM Module. (2012)
|
|           Generalized Linear Modeling
|
|           of
|
|           Multilevel (Hierarchical) Data
|
|
|
|
|           Copyright by Scientific Software International, Inc., 2012
|
|           Scientific Software International, Inc.
|
|           Website:  www.ssicentral.com
|
|           Support:  techsupport@ssicentral.com
|
|
|
|           DATE OF ANALYSIS: September 13, 2013
|
|           TIME OF ANALYSIS: 21H08:57
|
|           DATA FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
|           LISREL\DataImputada |
|           .LSF
|
|           INPUT FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
|           LISREL\DataImputada |
|           .PRL
|
|
|
o-----
--o
```

The following lines were read from the MAPGLIM syntax file:

```
MGLimOptions Converge=0.0001 MaxIter=100 MissingCode=9999
              Method=Quad NQUADPTS=10 ;
Title=Modelo 1;
SY='C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada.LSF';
ID2=SCHOOLID;
Weight=Peso;
Distribution=BER;
Link=PROBIT;
Intercept=Yes;
DepVar=XPCL50;
CoVars=Xgrado XSexo$ ZZONA$ ZTC$;
RANDOM2=intcept;
```

```

o=====o
| Modelo 1 |
|         |
o=====o

```

Model and Data Descriptions

Sampling Distribution = Bernoulli
 Link Function = Probit
 PROB(Success)= PHI(ETA)
 PHI denotes the CDF of the standard Normal distribution

Level-1 Weight Variable = Peso
 Number of Level-2 Units = 240
 Number of Level-1 Units = 5985
 Number of Level-1 Units per Level-2 Unit =

9	31	16	6	8	35	33	32	4	31	31	10
11	18	32	29	19	29	30	19	31	31	30	31
32	32	35	2	35	29	12	29	10	4	33	32
15	32	11	35	31	9	15	27	30	33	9	33
31	7	32	14	29	32	9	33	31	33	29	32
20	27	8	21	29	3	30	33	26	30	32	32
30	33	32	33	35	24	32	34	31	35	6	34
32	32	16	10	9	13	32	33	32	27	25	35
34	34	11	9	30	33	33	24	32	32	32	3
31	24	33	33	34	8	33	32	9	33	30	28
31	33	30	13	32	31	22	11	27	32	34	33
31	34	33	28	3	9	33	12	30	34	28	33
32	35	9	9	16	11	33	17	21	28	32	34
32	33	27	35	2	8	33	24	28	34	14	9
33	35	27	34	28	29	29	32	31	32	2	32
23	21	30	31	3	12	30	20	33	3	28	3
10	30	33	30	32	32	13	5	8	8	35	8
9	32	32	32	28	2	22	33	33	29	31	35
32	4	30	33	32	34	19	24	31	7	32	8
33	35	33	31	7	10	32	34	35	32	33	34

```

o=====o
| Descriptive statistics for all the variables in the model |
o=====o

```

Variable	Minimum	Maximum	Mean	Standard Deviation
XPCL5O1	0.0000	1.0000	0.9004	0.2995
XPCL5O2	0.0000	1.0000	0.0996	0.2995
intcept	1.0000	1.0000	1.0000	0.0000
Xgrado	7.0000	11.0000	9.8251	1.0177
XSexo1	0.0000	1.0000	0.4987	0.5000
ZZONA1	0.0000	1.0000	0.1272	0.3332
ZTC1	0.0000	1.0000	0.7866	0.4097

```

o=====o
| Results for the model without any random effects |
o=====o

```

Goodness of fit statistics

Statistic Ratio	Value	DF
---	-----	--
---	-----	--

Likelihood Ratio Chi-square	35552.7859	5979
5.9463		
Pearson Chi-square	847528.7148	5979
141.7509		

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept	-5.0680	0.3627	-13.9747	0.0000
Xgrado	0.4240	0.0351	12.0794	0.0000
XSexo1	0.2260	0.0501	4.5135	0.0000
ZZONA1	-0.6561	0.1324	-4.9573	0.0000
ZTC1	-0.9195	0.0498	-18.4576	0.0000

o=====o
 | Optimization Method: Adaptive Quadrature |
 o=====o

Number of quadrature points =	10
Number of free parameters =	6
Number of iterations used =	3

-2lnL (deviance statistic) =	2798.14251
Akaike Information Criterion	2810.14251
Schwarz Criterion	2850.32457

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept	-5.1356	0.4632	-11.0882	0.0000
Xgrado	0.4125	0.0437	9.4325	0.0000
XSexo1	0.1982	0.0628	3.1584	0.0016
ZZONA1	-0.9065	0.2392	-3.7902	0.0002
ZTC1	-1.0910	0.1335	-8.1703	0.0000

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
-----	-----	-----	-----	-----
intcept/intcept	0.4848	0.0813	5.9626	0.0000

Level 2 covariance matrix

	intcept
intcept	0.484754

Level 2 correlation matrix

	intcept
intcept	1.000000

Calculation of the intracluster correlation

residual variance = 1 (assumed)
cluster variance = 0.4848

intracluster correlation = $0.4848 / (0.4848 + 1.000) = 0.326$

o=====o
| LISREL Multilevel GLIM used 0.25 seconds CPU |
o=====o

ANEXO 5: Modelo 2 – Probit

```
o-----
--o
|
|           Student Edition of MGLIM Module. (2012)
|
|           Generalized Linear Modeling
|
|           of
|
|           Multilevel (Hierarchical) Data
|
|
|           Copyright by Scientific Software International, Inc., 2012
|
|           Scientific Software International, Inc.
|
|           Website:  www.ssicentral.com
|
|           Support:  techsupport@ssicentral.com
|
|
|           DATE OF ANALYSIS: September 13, 2013
|
|           TIME OF ANALYSIS: 21H20:12
|
|           DATA FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada |
|           .LSF
|
|           INPUT FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada |
|           .PRL
|
o-----
--o

The following lines were read from the MAPGLIM syntax file:
MGLimOptions Converge=0.0001 MaxIter=100 MissingCode=9999
          Method=Quad NQUADPTS=10 ;
Title=Modelo 2;
SY='C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada.LSF';
ID2=SCHOOLID;
Weight=Peso;
Distribution=BER;
Link=PROBIT;
Intercept=Yes;
DepVar=XPCL50;
CoVars=Xgrado XSexo$ XMETASUM XUNDREM XESCS XDISCLIM ZZONA$ ZTC$;
RANDOM2=intcept;

o=====o
| Modelo 2 |
o=====o
```

Model and Data Descriptions

Sampling Distribution = Bernoulli
 Link Function = Probit
 PROB(Success)= PHI(ETA)
 PHI denotes the CDF of the standard Normal distribution

Level-1 Weight Variable = Peso
 Number of Level-2 Units = 240
 Number of Level-1 Units = 5985
 Number of Level-1 Units per Level-2 Unit =

9	31	16	6	8	35	33	32	4	31	31	10
11	18	32	29	19	29	30	19	31	31	30	31
32	32	35	2	35	29	12	29	10	4	33	32
15	32	11	35	31	9	15	27	30	33	9	33
31	7	32	14	29	32	9	33	31	33	29	32
20	27	8	21	29	3	30	33	26	30	32	32
30	33	32	33	35	24	32	34	31	35	6	34
32	32	16	10	9	13	32	33	32	27	25	35
34	34	11	9	30	33	33	24	32	32	32	3
31	24	33	33	34	8	33	32	9	33	30	28
31	33	30	13	32	31	22	11	27	32	34	33
31	34	33	28	3	9	33	12	30	34	28	33
32	35	9	9	16	11	33	17	21	28	32	34
32	33	27	35	2	8	33	24	28	34	14	9
33	35	27	34	28	29	29	32	31	32	2	32
23	21	30	31	3	12	30	20	33	3	28	3
10	30	33	30	32	32	13	5	8	8	35	8
9	32	32	32	28	2	22	33	33	29	31	35
32	4	30	33	32	34	19	24	31	7	32	8
33	35	33	31	7	10	32	34	35	32	33	34

=====O
 | Descriptive statistics for all the variables in the model |
 =====O

Variable	Minimum	Maximum	Mean	Standard Deviation
XPCL501	0.0000	1.0000	0.9004	0.2995
XPCL502	0.0000	1.0000	0.0996	0.2995
intcept	1.0000	1.0000	1.0000	0.0000
Xgrado	7.0000	11.0000	9.8251	1.0177
XSexo1	0.0000	1.0000	0.4987	0.5000
XMETASUM	-2.0098	1.3443	-0.3356	0.9033
XUNDREM	-1.8824	1.4152	-0.4322	0.9311
XESCS	-4.8025	1.9933	-1.2962	1.2270
XDISCLIM	-2.8091	1.8379	0.1891	0.7686
ZZONA1	0.0000	1.0000	0.1272	0.3332
ZTC1	0.0000	1.0000	0.7866	0.4097

=====O
 | Results for the model without any random effects |
 =====O

Goodness of fit statistics

Statistic	Value	DF
Ratio		
---	----	--
Likelihood Ratio Chi-square	47958.7820	5975
8.0266		
Pearson Chi-square	42569476.7179	5975
7124.5986		

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-3.5324	0.4203	-8.4034	0.0000
Xgrado	0.2772	0.0407	6.8067	0.0000
XSexo1	0.1858	0.0559	3.3220	0.0009
XMETASUM	0.3807	0.0381	9.9899	0.0000
XUNDREM	0.2408	0.0322	7.4688	0.0000
XESCS	0.4052	0.0285	14.2356	0.0000
XDISCLIM	0.0358	0.0363	0.9858	0.3242
ZZONA1	-0.3663	0.1615	-2.2682	0.0233
ZTC1	-0.4945	0.0591	-8.3617	0.0000

o=====o
 | Optimization Method: Adaptive Quadrature |
 o=====o

Number of quadrature points = 10
 Number of free parameters = 10
 Number of iterations used = 3

 -2lnL (deviance statistic) = 2488.39419
 Akaike Information Criterion 2508.39419
 Schwarz Criterion 2575.36430

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-4.0300	0.4853	-8.3039	0.0000
Xgrado	0.3129	0.0463	6.7582	0.0000
XSexo1	0.1555	0.0655	2.3735	0.0176
XMETASUM	0.3850	0.0412	9.3508	0.0000
XUNDREM	0.2399	0.0347	6.9087	0.0000
XESCS	0.2878	0.0348	8.2705	0.0000
XDISCLIM	0.0439	0.0404	1.0873	0.2769
ZZONA1	-0.5174	0.2239	-2.3102	0.0209
ZTC1	-0.6913	0.1134	-6.0939	0.0000

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept/intcept	0.2564	0.0524	4.8907	0.0000

Calculation of the intraclass correlation

residual variance = 1 (assumed)
 cluster variance = 0.2564

intraclass correlation = 0.2564 / (0.2564 + 1.000) = 0.204

o=====o
 | LISREL Multilevel GLIM used 0.36 seconds CPU |
 o=====o

ANEXO 6: Modelo 3 – Probit

```
o-----
--o
|
|           Student Edition of MGLIM Module. (2012)
|
|           Generalized Linear Modeling
|
|           of
|
|           Multilevel (Hierarchical) Data
|
|
|           Copyright by Scientific Software International, Inc., 2012
|
|           Scientific Software International, Inc.
|
|           Website:  www.ssicentral.com
|
|           Support:  techsupport@ssicentral.com
|
|
|           DATE OF ANALYSIS: September 13, 2013
|
|           TIME OF ANALYSIS: 21H23:09
|
|           DATA FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada |
|           .LSF
|
|           INPUT FILE:
|
|           C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada |
|           .PRL
|
o-----
--o

The following lines were read from the MAPGLIM syntax file:
MGLimOptions Converge=0.0001 MaxIter=100 MissingCode=9999
          Method=Quad NQUADPTS=10 ;
Title=Modelo 3;
SY='C:\Users\user\Desktop\Avances 04-09-13\bd imputada final
LISREL\DataImputada.LSF';
ID2=SCHOOLID;
Weight=Peso;
Distribution=BER;
Link=PROBIT;
Intercept=Yes;
DepVar=XPCL50;
CoVars=Xgrado XSexo$ XMETASUM XUNDREM XESCS ZZONA$ ZTC$
        ZPROPQUA ZSCMATD;
RANDOM2=intcept;

o=====o
| Modelo 3 |
o=====o
```

Model and Data Descriptions

Sampling Distribution = Bernoulli
 Link Function = Probit
 PROB(Success)= PHI(ETA)
 PHI denotes the CDF of the standard Normal distribution

Level-1 Weight Variable = Peso
 Number of Level-2 Units = 240
 Number of Level-1 Units = 5985
 Number of Level-1 Units per Level-2 Unit =

9	31	16	6	8	35	33	32	4	31	31	10
11	18	32	29	19	29	30	19	31	31	30	31
32	32	35	2	35	29	12	29	10	4	33	32
15	32	11	35	31	9	15	27	30	33	9	33
31	7	32	14	29	32	9	33	31	33	29	32
20	27	8	21	29	3	30	33	26	30	32	32
30	33	32	33	35	24	32	34	31	35	6	34
32	32	16	10	9	13	32	33	32	27	25	35
34	34	11	9	30	33	33	24	32	32	32	3
31	24	33	33	34	8	33	32	9	33	30	28
31	33	30	13	32	31	22	11	27	32	34	33
31	34	33	28	3	9	33	12	30	34	28	33
32	35	9	9	16	11	33	17	21	28	32	34
32	33	27	35	2	8	33	24	28	34	14	9
33	35	27	34	28	29	29	32	31	32	2	32
23	21	30	31	3	12	30	20	33	3	28	3
10	30	33	30	32	32	13	5	8	8	35	8
9	32	32	32	28	2	22	33	33	29	31	35
32	4	30	33	32	34	19	24	31	7	32	8
33	35	33	31	7	10	32	34	35	32	33	34

=====o
 | Descriptive statistics for all the variables in the model |
 o=====o

Variable	Minimum	Maximum	Mean	Standard Deviation
-----	-----	-----	-----	-----
XPCL501	0.0000	1.0000	0.9004	0.2995
XPCL502	0.0000	1.0000	0.0996	0.2995
intcept	1.0000	1.0000	1.0000	0.0000
Xgrado	7.0000	11.0000	9.8251	1.0177
XSexo1	0.0000	1.0000	0.4987	0.5000
XMETASUM	-2.0098	1.3443	-0.3356	0.9033
XUNDREM	-1.8824	1.4152	-0.4322	0.9311
XESCS	-4.8025	1.9933	-1.2962	1.2270
ZZONA1	0.0000	1.0000	0.1272	0.3332
ZTC1	0.0000	1.0000	0.7866	0.4097
ZPROPQUA	0.0000	1.0000	0.4573	0.2812
ZSCMATED	-3.3887	1.9276	-1.1049	1.1895

=====o
 | Results for the model without any random effects |
 o=====o

Goodness of fit statistics

Statistic	Value	DF
Ratio	-----	--

Likelihood Ratio Chi-square	49552.0212	5974
8.2946		
Pearson Chi-square	93202584.1748	5974
15601.3700		

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-4.0359	0.4422	-9.1269	0.0000
Xgrado	0.2884	0.0419	6.8765	0.0000
XSexol	0.1824	0.0565	3.2273	0.0012
XMETASUM	0.3676	0.0387	9.4994	0.0000
XUNDREM	0.2344	0.0326	7.1970	0.0000
XESCS	0.3477	0.0294	11.8436	0.0000
ZZONA1	-0.2040	0.1614	-1.2645	0.2060
ZTC1	-0.2745	0.0659	-4.1636	0.0000
ZPROPQUA	0.5622	0.1113	5.0491	0.0000
ZSCMATED	0.1547	0.0274	5.6425	0.0000

o=====o
 | Optimization Method: Adaptive Quadrature |
 o=====o

Number of quadrature points = 10
 Number of free parameters = 11
 Number of iterations used = 3

 -2lnL (deviance statistic) = 2452.71490
 Akaike Information Criterion 2474.71490
 Schwarz Criterion 2548.38203

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-4.4800	0.5043	-8.8836	0.0000
Xgrado	0.3176	0.0466	6.8191	0.0000
XSexol	0.1652	0.0648	2.5490	0.0108
XMETASUM	0.3770	0.0412	9.1558	0.0000
XUNDREM	0.2393	0.0346	6.9180	0.0000
XESCS	0.2649	0.0345	7.6736	0.0000
ZZONA1	-0.2902	0.2131	-1.3621	0.1732
ZTC1	-0.3585	0.1145	-3.1302	0.0017
ZPROPQUA	0.6610	0.1807	3.6579	0.0003
ZSCMATED	0.2028	0.0465	4.3644	0.0000

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept/intcept	0.1882	0.0411	4.5751	0.0000

Calculation of the intraclass correlation

residual variance = 1 (assumed)
 cluster variance = 0.1882

intraclass correlation = 0.1882 / (0.1882 + 1.000) = 0.158

o=====o
 | LISREL Multilevel GLIM used 0.41 seconds CPU |
 o=====o